USING EARNED VALUE DATA TO FORECAST THE DURATION OF DEPARTMENT OF DEFENSE (DOD) SPACE ACQUISITION PROGRAMS

THESIS

Shedrick M. Bridgeforth, First Lieutenant, USAF

AFIT-ENV-MS-15-M-177

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-15-M-177

USING EARNED VALUE DATA TO FORECAST THE DURATION OF DEPARTMENT OF DEFENSE (DOD) SPACE ACQUISITION PROGRAMS

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Shedrick M. Bridgeforth, MBA, BS

First Lieutenant, USAF

March 2015

AFIT-ENV-MS-15-M-177

USING EARNED VALUE DATA TO FORECAST THE DURATION OF DEPARTMENT OF DEFENSE (DOD) SPACE ACQUISITION PROGRAMS

Shedrick M. Bridgeforth, MBA, BS

First Lieutenant, USAF

Committee Membership:

Lt Col Jonathan D. Ritschel, Ph.D
Chair

Dr. Edward D. White, Ph.D
Member

Capt Grant Keaton
Member

AFIT-ENV-MS-15-M-177

Abstract

The accuracy of cost estimates is vital during this era of budget constraints. A key component of this accuracy is regularly updating the cost estimate at completion (EAC). A 2014 study by the Air Force Cost Analysis Agency (AFCAA) improved the accuracy of the cost estimate at completion (EAC) for space system contracts. The study found schedule duration to be a cost driver, but assumed the underlying duration estimate was accurate. This research attempts to improve the accuracy of the duration estimate from the AFCAA study. First, the overall accuracy is evaluated with the Mean Absolute Percent Error (MAPE). Then the duration estimates are analyzed for timeliness to determine when the methods offer improved accuracy over the status quo. Finally, the methods are evaluated for reliability (accuracy for contracts with Over-Target-Baselines (OTBs)). The methods researched here are more accurate, timely, and reliable than the status quo method. The original objective, to improve the accuracy of the duration estimates for the cost estimating model, was achieved. The accuracy gains ranged from 2.0% to 13.4% for single contracts, 3.2% to 5.1% for OTB contracts, and 2.9% to 5.2% for all contracts combined. The accuracy improvement is more pronounced from 0% to 70% completion, with a 4.0% to 7.6% increase in accuracy. Finally, the overall accuracy improvement for the EAC was 6.5% (24.4% vs. 17.9%).

This thesis is dedicated to my lovely wife. Its completion would not be possible without her support during the countless hours researching, analyzing, and writing.

Acknowledgments

I would like to express my gratitude to my thesis committee. First and foremost, I would like to thank my faculty advisor, Lt Col Ritschel; your patience, insight, and guidance made this research effort less treacherous. I would also like to thank Dr. White for sharing his statistical expertise. Finally, I would like to thank my sponsor, Capt Grant Keaton, for providing the thesis topic and supporting this research effort.

Shedrick M. Bridgeforth

Table of Contents

List of Figures

List of Tables

Page

Using Earned Value Data to Forecast the Duration of Department of Defense (DoD) Space Acquisition Programs

I. Introduction

General Issue

The Department of Defense (DoD) faces a constrained fiscal environment for the foreseeable future. Under these conditions, the DoD has come under increased scrutiny from Congress to improve the accuracy of estimating acquisition programs' cost and schedule. Many prior studies have focused on the overall cost of programs (the cost estimate at completion (EAC)) (Smoker, 2011). However, cost is not the only important measure of performance. Cost, schedule, and technical performance are the three primary performance objectives of acquisition program management. These three components are inter-related, therefore when one component is affected, the others are affected. Although cost performance is studied, schedule performance is the primary focus of this research with an emphasis on improving the accuracy of schedule estimates.

The current method for evaluating schedule performance is based on Earned Value Management (EVM), an approach created in the 1960s. EVM has been a useful tool for monitoring cost performance, but it has limitations with assessing schedule performance (Lipke, 2003). Specifically the schedule performance index (SPI) indicates whether a contract's schedule performance is favorable (SPI > 1.0) or unfavorable (SPI < 1.0). Unfortunately, the SPI converges to 1.0 as the contract nears completion; as the contract matures the SPI gradually becomes useless as a schedule performance metric. Earned Schedule (ES), a schedule performance metric, was developed to overcome

EVM's shortcomings (Lipke, 2003). Earned Schedule has demonstrated improved schedule performance assessment over SPI (Henderson, 2004; Crumrine, 2013). However, Earned Schedule has not been applied exclusively to estimating the duration of space system acquisitions. This research explores and applies five techniques to estimate the duration at completion for space programs. The objective is to enhance cost estimates and decision support. This chapter provides a discussion of how schedules are estimated and evaluated with an overview of EVM based methods and the critical path method (CPM). The remainder of the chapter will address the specific research questions to be investigated, methodology used, and the limitations of this research.

Background

The traditional project control method (EVM) monitors actual performance compared to planned, analyzes the variance, and provides a quantitative method to forecasts the end result (Abdel Azeem, Hosny, & Ibrahim, 2014). Research conducted by the Air Force Cost Analysis Agency (AFCAA) revealed EVM estimating methods improved cost estimates of space systems midway through the acquisition lifecycle (Keaton, 2014). A key component of that study was the use of duration as a cost driver (Keaton, 2014). However, one potentially problematic assumption of that study was the assumption of accuracy for the duration estimates. The duration estimates were based on the contractor performance reports (CPR) which are based on the critical path method (CPM). Are the CPR duration estimates accurate for space systems? The simple answer is no. Schedule growth is rampant in DoD acquisition; satellite programs experience above average development cost and schedule growth (GAO, 2014). Why does schedule growth occur? According to a recent RAND report, *Prolonged Cycle Times and Schedule Growth in Defense Acquisitions*, the top three cited factors for schedule were:

- Difficulty in managing technological risk
- Overoptimistic initial estimates and expectations
- Lack of funding stability (2014)

These factors can be grouped into two categories: errors and decisions. *Errors* include cost estimation, schedule estimation, and technical issues (development or implementation) (Bolten, et al., 2008). *Decisions* include changes in requirements, affordability, quantity, schedule, and funding transfers (within or between a program) (Bolten, et al., 2008). Even perfect estimates cannot account for all of the impacts from decisions. Therefore the CPR estimates may not be accurate at all times. On the other hand, in the absence of decision effects, the CPR estimates may not be accurate due to overoptimistic expectations. Why use the CPR based duration estimates? One reason is a lack of better alternatives. Given these shortcomings, the opportunity exists to provide a more accurate duration estimate.

Problem Statement

Cost estimates play a vital role in the budgeting process. Historically, schedule estimates are not given the same level of attention as cost estimates (GAO, 2012). However, schedule estimates are also essential to the accuracy of cost estimates and overall program performance (GAO, 2012). The accuracy of a cost estimate is important because a lack of accuracy has unfavorable consequences. Cost estimates that underestimate may eventually require funds to be pulled from other programs causing extra work, loss of productivity, and possibly jeopardizing multiple programs (Bolten, et al., 2008). Overestimating may lead to an opportunity cost; resources that could have been allocated to systems were not invested. Ultimately, more accurate cost estimates will lead to better resource allocation decisions and inputs into the budget process.

Since 1993 there have been many studies utilizing earned value data to develop cost estimates (Christensen, 1993, 1994, 1999; Unger, 2001; Nystrom, 1995). These studies employed a variety of methods: index-based, linear regression, nonlinear regression, and S-curves. The overwhelming result of these studies is there is not one method that works best in all circumstances (Trahan, 2009). The AFCAA study determined Estimates at Completion (EACs) based on the Budgeted Cost of Work Performed (BCWP) burn rate improved the accuracy for space systems with developmental contracts (Keaton, 2014). The question remained, are the underlying duration estimates accurate? This research attempts to evaluate the schedule estimating method used in the AFCAA study. Next, additional methods are explored in an effort to improve the accuracy.

In addition to cost estimate problems, the majority of space programs have schedule growth (Younossi, et al., 2008). Therefore, a need exists to accurately predict program duration in order to detect schedule issues sooner. Improved schedule forecasts should provide more accurate and timely data to program managers thus enhancing risk management and decision making.

The current methods (CPM and EVM) for estimating program duration are adequate, but can be improved. Many studies explain the strengths and weakness of traditional EVM (Lipke, Zwikael, Henderson, & Anbari, 2009) and CPM (Kim, 2007).

The primary weaknesses of the CPM are failure to update the estimate with actual data, the lack of early detection of schedule problems, and complexity (GAO, 2012). The foundation of the argument against EVM is that it is value based instead of time based and deterministic instead of probabilistic (Lipke, 2003; Kim, 2007). For example, a schedule variance (Earned Value - Planned Value) of $3M means we are behind schedule $3M instead of three months behind.

Earned Schedule was developed to overcome the value based weakness of EVM. However, both EVM and Earned Schedule forecasts only provide point estimates so they do not provide a probability or uncertainty associated with the estimate. The Kalman filter earned value method (KEVM) addresses the inherent weaknesses of CPM, EVM, and ES (Kim & Reinschmidt, 2010). This method is a hybrid of earned schedule (ES) and a Kalman filter and has shown improved accuracy over the current methods (CPM, EVM, and ES) (Kim & Reinschmidt, 2010). This research will not attempt to replace EVM techniques. Instead, the research objective is to enhance and expand the toolset for estimating program duration.

Research Objective and Questions

The overall research objective is to evaluate forecasting methods for space program duration based on the following criteria: accuracy, reliability, and timeliness. In support of the overarching research objective, the following questions will be investigated:

1. What are the appropriate methods to estimate a program's duration?
2. How should accuracy be measured and how accurate are the various schedule estimating methods (individual contract, overall and by various groupings)?

3. At what point in time (if at all) are the new techniques more accurate than the status quo?
4. Are the forecasts accurate for programs with one or more over target baseline (OTB)?

The overall goal of this research is to determine the schedule estimating methods that can improve the cost estimate and add value to space system program offices (SPOs). This value may be in the form of an additional tool for analysts to use when evaluating the schedule performance of a program. The first investigative question addresses what forecasting methods are available. The second investigative question is twofold; first we must determine which accuracy measure should be used. Then we must analyze the accuracy of each method by individual contract, overall, and groupings to determine if substantial difference exist in the forecasting models. The third investigative question seeks an answer to when, if at all, the forecasting methods become more accurate than the status quo. Generally earlier forecasts are less accurate because more uncertainty exists. Additionally, most programs are not stable until later in the program (50% complete or later) and developmental programs take longer to stabilize than production programs (Petter, 2014). The fourth question determines whether the forecasts are still useful for programs that have OTBs. Many programs have undergone an OTB. Programs that undergo an OTB may be less stable than non-OTB programs.

Methodology

The Defense Cost and Resource Center (DCARC) is used to obtain the necessary EVM data to conduct the analysis of program schedule. This research will examine forecasts based on the critical path method (CPM), earned value and earned schedule index based methods, time series, regression (Smoker, 2011), the Kalman Filter Forecasting Method (Kim, 2007), and analysis of the Integrated Master Schedule (IMS). All of the forecasting methods will use data from the Earned Value Management Central Repository (EVM-CR).

The accuracy of the models will be evaluated by the mean absolute percentage error (MAPE). The goal is to measure the overall accuracy of each model and the accuracy at certain percent complete intervals: 0-10%, 11-20%, and so on until 100%. The forecasting methods will first be evaluated by individual contract. Then the contracts are aggregated by duration: long, medium, and short duration. Next the contracts are grouped by OTBs (one or more) and non OTB contracts. Finally, accuracy is evaluated across all contracts (all observations).

Assumptions and Limitations

The DCARC is a system to collect Major Defense Acquisition Program (MDAP) data (DCARC, 2014). These data consist of Contactor Performance Reports (CPR) and other information needed to evaluate program performance. The primary EVM data of interest in this research are: Budgeted Cost of Work Performed (BCWP), Budget at Complete (BAC), program start date, and the estimated completion date (ECD) for the program. The government contractors required to provide CPRs must adhere to industry standards for EVM systems and reporting. The CPR data is reviewed by the program management office for its quality and completeness. Although no data source is without error, the DCARC is assumed to be a credible and reliable data source because of the industry standard in place and the program office review process (NDIA/PMSC, 2012). As an added check, we reviewed the CPR data used in this research for accuracy, completeness, and consistency.

The analysis database is limited to space system programs primarily because the characteristics of space systems programs are different than other programs such as aircraft. Typically, space systems are acquired in much lower quantities than other programs. Strictly analyzing space systems should lead to a more accurate approach for estimating space systems, but could be less useful for other systems. The specific type of contract selected for this analysis is Research, Development, Test and Evaluation (RDT&E). RDT&E programs are more susceptible to schedule and cost estimating errors than production contracts (Bolten, et al., 2008). This result is logical because production contracts are for more mature programs with less uncertainty than development contracts (Bolten, et al., 2008; Keaton, 2014). Therefore in theory, RDT&E schedule estimates have more room for improvement.

Thesis Preview

A program's schedule is important because programs completed on time will deliver capability sooner. Additionally, schedule is important because of its relationship with cost. Generally, schedule delays lead to increased program costs because extra resources and/or overtime are utilized to reduce the delay (GAO, 2012). This research does not attempt to study the underlying causes of schedule delays. Rather, this research attempts to forecast the duration of individual contracts based on actual data.

One critical component of cost analysis is to reduce risk by regularly updating cost estimates as programs mature (GAO, 2009; Keaton, 2014). Keaton's study demonstrated improved accuracy with cost estimates using duration as a parameter in the following equation (2014):

Equation 1: Estimate at Complete ($\mathrm{EAC_{BCWP}}$)

$$\mathrm{EAC_{BCWP}} = \left(\mathrm{Month_{Est\ Completion}} - \mathrm{Month_{Current}}\right) * \mathrm{BCWP_{Burn\ Rate}} + \mathrm{BCWP_{To\ Date}}$$

Where the $\mathrm{BCWP_{Burn\ Rate}}$ is calculated via linear regression with BCWP as the dependent variable and time (months) as the independent variable. The key relationship is the time to complete the system and the burn rate. Therefore, increasing the accuracy of the underlying duration estimate should further improve the accuracy of the BCWP based cost estimate (Equation 1).

Chapter 2 examines the relevant literature for program management, EVM, Earned Schedule (ES), and the Critical Path Method (CPM). Additionally, two established forecasting techniques are described: time series analysis and the Kalman filter method. Finally, we examine a new technique to forecast a contract's schedule based on the Integrated Master Schedule (IMS). Chapter 3 discusses the specific methodology used in this research. Chapter 4 presents the results of the research and a detailed discussion. Chapter 5 summarizes the research, discusses the recommendations, and explores areas for future research.

II. Literature Review

Chapter Overview

The purpose of this chapter is to research program management, EVM, and forecasting literature in order to develop accurate duration estimates. The first objective is to explain program management and EVM in further detail. Then schedule forecasting techniques are described, which leads into the relevant EVM research and the emergence of Earned Schedule. Next, linear regression, time series analysis, Kalman filter theory, and the Kalman filter forecasting method are examined. Finally, an analysis of the Integrated Master Schedule (IMS) is presented.

Program Management

Fleming and Koppelman define a project as "a one-time-only endeavor to achieve specific objectives with a precise start and completion date and finite resources to accomplish the goals." (2000: 203) Whereas a program is essentially a portfolio of two or more related projects (Peisach & Kroecker, 2008). The literature often uses project and program management interchangeably. This research will stay consistent with the previous definitions. Individual contracts are considered projects. Program will be used when discussing the overall performance of the portfolio of contracts.

According to the GAO, "[the] DoD and Congress have taken meaningful steps to improve the acquisition of major weapon systems, yet many programs are still falling short of cost and schedule estimates" (GAO, 2014: 1). Program managers are responsible for the overall success of the program based on three primary criteria: cost, schedule, and technical performance. In order to monitor a program's performance, the *Defense*

Acquisition Guidebook states, "the program manager should obtain integrated cost and schedule performance data at an appropriate level of summarization to monitor program execution" (2014). Earned Value Management is the DoD's primary method for project/program execution and control. The EVM approach can be used to monitor and evaluate cost and schedule performance while attempting to meet technical objectives.

Earned Value Management Background

Earned Value Management (EVM) is an industry best practice for program management and is mandatory for large DoD acquisition programs (GAO, 2009). EVM goes further than a simple comparison of budgeted costs to actual costs. The budgeted cost of work scheduled (planned value), the budgeted cost of work performed (earned value), and the actual cost of work performed (actual value) are used to develop performance metrics. These metrics can then be used to assess the program's cost and schedule performance and to estimate cost and time to complete (GAO, 2009). The *Defense Acquisition Guidebook* defines EVM as:

> A key integrating process in the management and oversight of acquisition programs, to include information technology projects… [and is an] approach that has evolved from combining both government management requirements and industry best practices to ensure the total integration of cost, schedule, and work scope aspects of the program. (Defense Acquisition University, 2014: 11.3.1)

Government acquisition programs exceeding a $20M budget must adhere to EVM standards (Defense Acquisition University (DAU), 2014). Programs over $50M must adhere to EVM standards and have a Defense Contract Management Agency (DCMA) validated EVM system (Defense Acquisition University (DAU), 2014). Figure 1 depicts the integration of program management, EVM, cost analysis, and systems engineering (GAO, 2009). Of specific importance is how cost analysis and cost estimates support the EVM process while program management monitors the entire process.

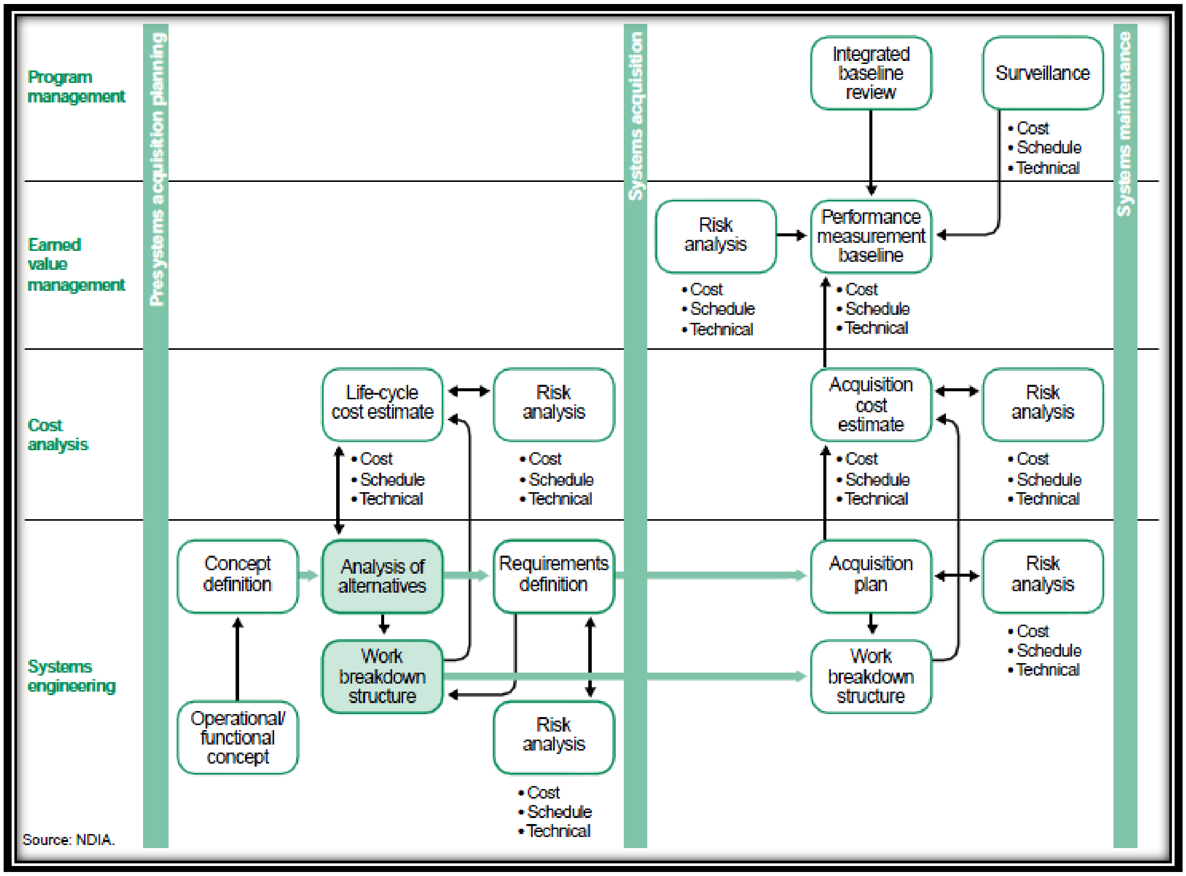


Figure 1: Cost Estimation, Systems Development, and Risk Management

Earned Value Management Data

The three fundamental EV data for assessing program performance are the Budgeted Cost of Work Scheduled (BCWS), the Budgeted Cost of Work Performed (BCWP), and the Actual Cost of Work Performed (ACWP). The contractor must report the data on a regular basis, usually monthly. The data are reviewed by the program management office before being entered into the Defense Cost and Resource Center (DCARC) database and the EVM-Central Repository. Table 1 summarizes and describes the relevant data available in the EVM Central Repository (EVM-CR) while Table 2 lists common metrics and formulas (DAU Gold Card, 2014). The primary EVM data of interest for schedule assessment are: the BCWP, BCWS, Budget at Completion (BAC), Start Date, and the Estimated Completion Date (ECD). These data are used to calculate many of the metrics in Table 2 and are the foundation for the duration forecasts. The duration forecast approach used in the AFCAA study is discussed in the next section (Keaton, 2014).

Table 1: Summary of EVM Measurements

EVM measurement	**Description**
Budgeted Cost of Work Scheduled (BCWS), also called Planned Value (PV)	Time-phased Budget Plan for work currently scheduled
Budgeted Cost of Work Performed (BCWP), also called Earned Value (EV)	Value of completed work in terms of the work's assigned budget
Actual Cost of Work Performed (ACWP), also called Actual Cost (AC)	Cost actually incurred in accomplishing work performed
Budget at Completion (BAC)	The planned total cost of the contract
Report From	The first day of the current reporting period for the contractor performance report (CPR)
Start Date	The date the contractor was authorized to start work on the contract, regardless of the date of contract definitization.
Completion Date	The completion date to which the budgets allocated in the PMB have been planned. This date represents the planned completion of all significant effort on the contract. The cost associated with the schedule from which this date is taken is the Total Allocated Budget.
Estimated Completion Date (ECD)	The contractor's latest revised estimated completion date. This date represents the estimated completion of all significant effort on the contract. The cost associated with the schedule from which this date is taken is the "most likely" management EAC.
Budget Completion Date	The contract scheduled completion date in accordance with the latest contract modification. The cost associated with the schedule from which this date is taken is the Contract Budget Base.

Table 2: EVM Metrics and Formulas

EVM measurement	Description	Formula
Cost Variance (CV)	Difference between planned and actual cost accomplishment	BCWP - ACWP
Schedule Variance (SV)	Difference between planned and actual schedule accomplishment, in dollar amount	BCWP - BCWS
Cost Performance Index (CPI)	Cost efficiency of a program	BCWP / ACWP
Schedule Performance Index (SPI)	Schedule efficiency of a program	BCWP / BCWS
Budgeted Cost for Work Remaining (BCWR)	The budgeted cost of uncompleted work packages to reach program's completion	BAC - BCWP
Estimate at Completion (EAC)	Forecasted total cost of program	[(BAC - BCWP) / PF] PF = CPI or SPI*CPI
Percent Complete (PC)	Percentage of the entire program that is complete	BCWP / BAC
To Complete Performance Index (TCPI)	Projects what the CPI will be for the remainder of the project to meet the BAC	[(BAC-BCWP) / (Target-ACWP)] Target = BAC, LRE, or EAC
Baseline Execution Index (BEI)	How well the project is following the baseline plan and completing baseline tasks as they are scheduled to be completed	[Total Baseline Tasks Completed / Total Tasks with Baseline Finish On or Prior to Current Report Period]

Schedule Forecasting: Critical Path Method

The *GAO Schedule Assessment Guide* defines the critical path as "the path of longest duration through the sequence of activities" (GAO, 2012: 4). Any delayed activities on the critical path will delay the entire project and therefore increase the project's duration (Fleming & Koppelman, 2000). The current DoD best practice for estimating program duration is the critical path method (CPM) in conjunction with the integrated master schedule (IMS). In addition to identifying important activities, the CPM is used to estimate the duration of the program (the reported ECD).

The *GAO Schedule Assessment Guide* considers updating the IMS with actual progress as a best practice for the CPM (2012). Unfortunately, that same report lists multiple occasions where programs failed to update the IMS (GAO, 2012). Given this shortcoming, the IMS alone may not be a sufficient schedule forecasting tool. For an MDAP, thousands of tasks are entered into the baseline schedule; additional tasks are added as the program matures further adding to the schedule's complexity. Because of this phenomenon, Lipke et al. argue that an "in depth schedule analysis is burdensome and may have a disruptive effect on the project team." (2009: 407). A less arduous method than an in depth schedule analysis is needed. However, this alternate approach must be at least as accurate as the CPM. Previous project schedule research has attempted to improve schedule forecasting using EVM data. This research will attempt to improve schedule forecasting over the CPM while remaining accessible (not overly complex or burdensome).

Schedule Forecasting: Earned Value Based Methods

The cancellation of the Navy's A-12 Avenger program in 1991 ignited a renewed interest in EVM research. These studies were focused on independent cost estimates at complete (IEAC) and they established EVM as an effective tool for estimating a program's cost performance (Christensen, 1993, 1994, & 1999). However, EVM's ability to forecast schedule has not been as successful. Henderson studied EVM based schedule forecasting with the three following formulas (2004):

Equation 2: Independent Estimate at Complete (IEAC)

$$IEAC(t) = PD / SPI$$

Equation 3: Independent Estimate at Complete (IEAC)

IEAC(t) = PD / SPI(t)

Equation 4: Independent Estimate at Complete (IEAC)

IEAC(t) = PD / CPI*SPI(t)

Where PD is the planned duration and SPI(t) is the earned schedule application to the SPI developed by Lipke. Equation 3 was the only accurate forecasting method out of the three in Henderson's study (2004). A potential weakness of this study is its application to only two projects: Commercial IT Infrastructure Expansion Project (Phase 1 and combined Phases 2 and 3) with durations of 34 and 22 weeks. The durations of these projects are short when compared to the duration of the space systems researched in this thesis (from 25 to 242 months). On the other hand, Henderson's method should be robust because it incorporates the CPM derived Planned Duration (PD) and EVM based Performance Factors (PF). Because of its robustness and simplicity, Henderson's basic formula [IEAC(t) = PD/ Performance Factor (PF)] is used as one of the primary forecasting methods in this research.

EVM research by Kim used the following formula to calculate an IEAC(t) he called the Estimated Duration at Completion (EDAC) (2007):

Equation 5: Estimated Duration at Completion (EDAC):

$$\mathrm{EDAC} = \text{time elapsed} + \frac{(\mathrm{BAC} - \mathrm{EV})}{\mathrm{SPI}}$$

Kim provided an example of a 120 month project to illustrate the schedule forecasting weakness of SPI (2007). Figure 2 shows the planned value (BCWP), the actual costs (ACWP), and the earned value (BCWS) over time intervals for this project (Kim, 2007).

The project has a 20% overrun in cost and schedule. Figure 3 shows the stable cost estimate at complete (EAC) and the erratic behavior of the Estimated Duration at Completion (EDAC) (Kim, 2007). The EDAC is overestimated by as much as 58% during the first half of the project. Furthermore, the EDAC is underestimated by 20% late in the project (95 months). This erratic behavior by the SPI based schedule forecast is also demonstrated in Henderson's research. However, the project examined in Kim's study is not described and the proposed equation does not match other schedule estimating formulas in the literature (2007). Therefore the results may not be generalizable.

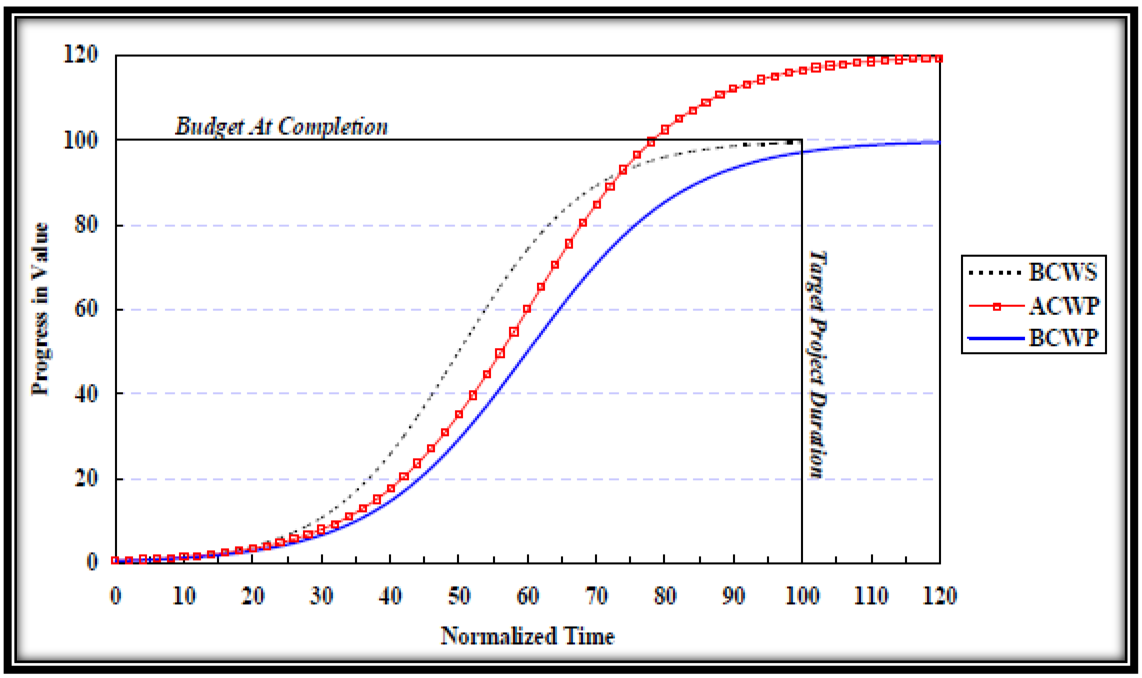


Figure 2: EVM Measurements over Time

To overcome the SPI schedule forecasting weakness, Lipke introduced the concept of Earned Schedule (2003). Earned Schedule is calculated as the number of time periods (N) earned value (BCWP) exceeds planned value (BCWS) plus a fraction of the earned value into the next period. Essentially, Earned Schedule is a linear interpolation of the Program Management Baseline (PMB) which is illustrated in Figure 4 as the Planned Value line (Lipke, 2012).

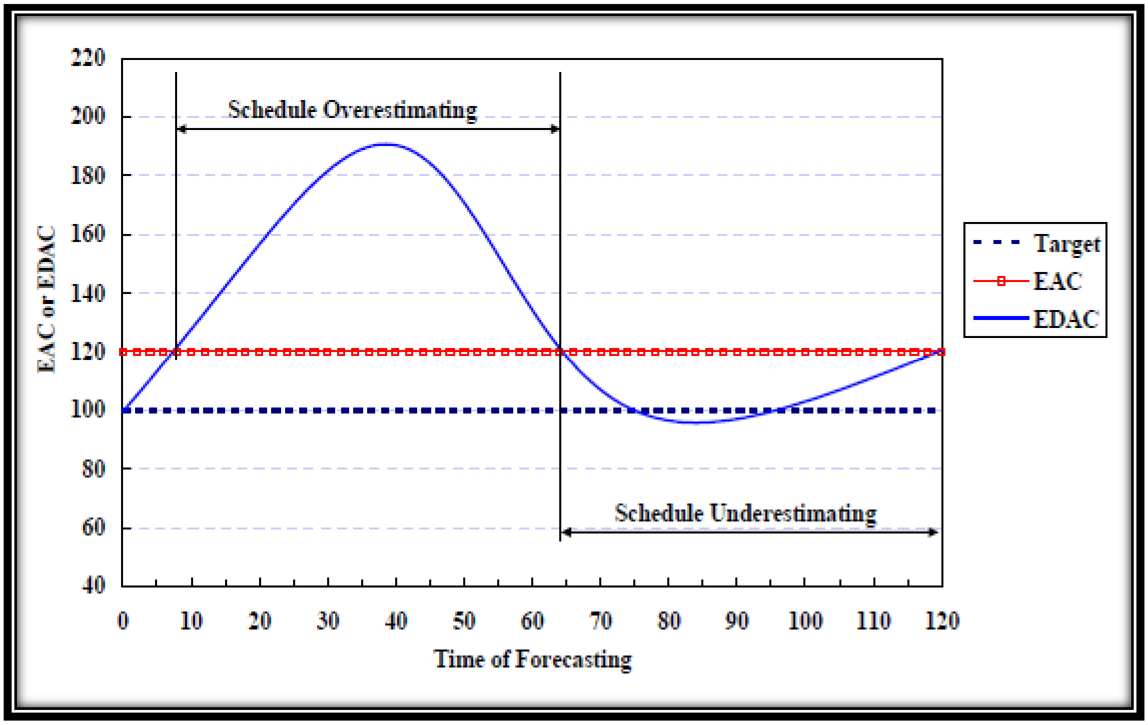


Figure 3: EAC and EDAC over Time

Lipke's Earned Schedule is calculated with the following equation (2012):

Equation 6: Earned Schedule

$$\text{Earned Schedule} = \text{N} + \frac{\text{Earned Value(current)} - \text{Planned Value(previous)}}{\text{Planned Value(current)} - \text{Planned Value(previous)}}$$

The Schedule Performance Index (SPI(t)) calculation is shown in Equation 7 (Lipke, 2012).

Equation 7: SPI(t)

$$\text{SPI(t)} = \frac{\text{Earned Schedule}}{\text{AT (actual time elapsed)}}$$

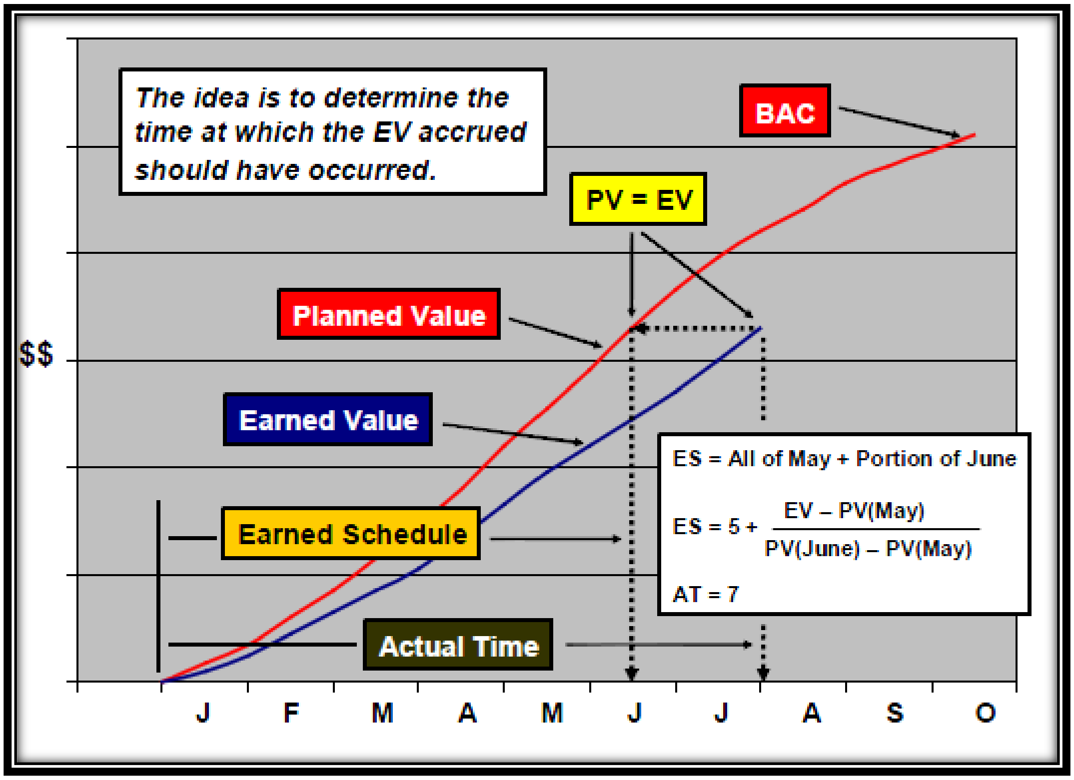


Figure 4: Earned Schedule Concept

Earned Schedule was originally developed to provide more sensible information to program managers (units of time instead of dollars). However, Henderson's study established SPI(t) as a useful forecasting method. Lipke et al. (2009) enhanced the SPI(t) forecasts by adding confidence intervals. That study applied a statistical approach to twelve projects and demonstrated accurate results for the three completion points (10%, 30%, and 60%). However, the projects used in the analysis were small (less than $6 million budget) and the specific projects types were not discussed.

Cost estimating methods were more numerous in the literature. Table 3 displays methods to forecast the cost estimate at completion where the base equation (EAC = time now + [(BAC - EV) / PF]) is similar to Henderson's Equation 5 (Anbari, 2003; Christensen, 1993; Lipke, 2003). This research will use some of the performance factors

(PF) from Table 3 to develop time estimates at completion (TEAC). The performance factors will be used with the planned duration (TEAC = PD/PF).

Table 3: Formulas for the Estimate at Completion (EAC)

Type	Performance Factor	Description
Standard	PF = SPI	Standard SPI
Earned Schedule	PF = SPI(t)	Earned Schedule based SPI
Schedule Cost Index	PF = CPI*SPI	The product of CPI and SPI is called the critical ratio (Anbari, 2003) or the Schedule Cost Index (Christensen, 1993).
Moving Average	PF = CPI(m)	Moving average of incremental CPI over latest month (m) intervals. For example: CPI(3m), CPI(6m), and CPI(12m).
% Complete	PF = (PC)*CPI+(1-PC)*SPI	A weighted method using percent complete (PC), CPI, and SPI

Vandevoorde and Vanhoucke (2006) examined three schedule forecasting model summarized in Table 4. That study used data from three projects at Fabricom Airport Systems in Brussels; the authors found earned schedule method as the only method with reliable results during the entire project (Vandevoorde & Vanhoucke, 2006: 298).

Table 4: Three Schedule Forecasting Methods

Type	EDAC	Description
Planned Value Method	EDAC = PD/PF [PF = SPI or SCI]	PD = planned duration
Earned Duration Method	EDAC = t + PD-ED/PF [PF = SPI or SCI]	ED = earned duration, [ED = actual duration*SPI]
Earned Schedule Method	EDAC = t + PD-ES/PF [PF = SPI(t)]	SPI(t) = ES/actual time

In 2011, Earned Schedule was studied by an AFIT student, Captain Kevin Crumrine. This study established the Earned Schedule based SPI(t) as a better indicator than SPI for assessing a program's schedule performance. Crumrine's study was focused on predicting schedule overruns of aircraft and missile systems rather than forecasting duration. However, it may provide insight into which Performance Factor (PF) leads to a better forecast. Because the SPI converges to 1.0 at approximately the 66 percent completion point of the program it may lose forecast accuracy as the program matures (Crumrine, 2011).

Earned Schedule appears to be the best EV based schedule forecasting method based on studies conducted by Henderson (2003), Lipke (2003), Lipke et al (2009), Vanhoucke & Vandevoorde (2006), and Crumrine (2011). With the exception of Crumrine, those studies focused on small acquisition programs and construction projects. A study forecasting the duration of space programs with EV data has not been conducted. This research attempts to fill that void in the literature.

Schedule Forecasting: Linear Regression

Linear regression has also been used to forecast a program's duration. A study by Smoker demonstrated this technique by first regressing the BCWP against months and the same approach for BAC (2011). In that study, Smoker set the BCWP intercept to zero because at the start of the project (time zero) the BCWP is zero. With the regression equations for BCWP and BAC, the next step is setting BCWP equal to BAC to solve for the unknown month as displayed in Equation 8. An assumption of this technique is the program is 100% complete when BCWP/BAC = 1.0 (Smoker, 2011). After the intermediate calculation, the duration formula is simplified to Equation 9.

Equation 8: Intermediate Calculation

$$\text{BCWP coefficient} * \text{Months} = \text{BAC intercept} + \text{BAC coefficient} * \text{Months}$$

Equation 9: Duration Forecast (Regression Based)

$$\text{Months} = \frac{\text{BAC intercept}}{(\text{BCWP coefficient} - \text{BAC coefficient})}$$

The primary strength of this method is it takes BAC growth into account; this may lead to better forecasts because it is attempting to predict the completion date based on trends instead of relying on the static reported completion date. Even in stable programs the BAC tends to gradually increase until the program nears completion. However, in unstable programs not only does the BAC gradually increase, the BAC also jumps from one reporting period to the next and exhibits a stepped relationship instead of a straight line. Because of this phenomenon, this regression based method may not be a useful forecasting approach for unstable programs. Another concern with this study is the lack of transparency in the program analyzed. This analysis was conducted on one program which was not described by name, commodity, or contract type. Furthermore, the early and late forecasts may not be accurate because the assumption of linearity occurs from approximately the 25% to 80% complete points. Finally, this method requires a basic understanding of linear regression and/or the software to conduct the regression.

Schedule Forecasting: Time Series Analysis

According to Box, Jenkins, and Reinsel, “a time series is a sequence of observations taken sequentially in time” (2008: 1). EVM data are reported on a monthly basis therefore they can be categorized as time series data. A key feature of a time series is that future observations are dependent on previous observations (Box, Jenkins, & Reinsel, 2008). Time series analysis is concerned with measuring dependence, building statistical models, and applying the models to important areas (Box, Jenkins, & Reinsel, 2008). These areas include: meteorology, economics, marketing, production, logistics, and financial markets (Makridakis, Wheelwright, & Hyndman, 1998). This research uses time series analysis to forecast future EV indices (CPI, SPI, SPI(t), and BEI) with past observations.

Forecasting with Time Series

Makridakis, Wheelwright, and Hyndman define forecasting as "the prediction of values of a variable based on known or past values of that variable or other related variables" (1998: 599). The basic forecasting process is an analysis of the data series and selection of the model that best fits the data series (Makridakis, Wheelwright, & Hyndman, 1998). There are many forecasting methods ranging from simple to complex; these methods include simple moving averages, exponetial smoothing, linear regression, general ARIMA, and seasonal ARIMA models. This research focuses on the Box-Jenkins method to building forecasting models.

Box Jenkins

Autoregressive (AR) / Integrated (I) / Moving Average (MA) (ARIMA) models were popularized by George Box and Gwilym Jenkins in the 1970s (Makridakis, Wheelwright, & Hyndman, 1998). The overall approach to building ARIMA models is called the Box-Jenkins methodology. The methodology contains three phases: identification, estimation and testing, and application (Makridakis, Wheelwright, & Hyndman, 1998). The major advantage to the Box-Jenkins approach is the robust evaluation of the underlying pattern of the time series baseline. The type of pattern that exists helps the practitioner decide which techniques to implement. Certain patterns suggest the data are suitable for AR, MA, I, or a combination of the parameters. The underlying statistical concepts are discussed in the subsequent sections followed by a discussion of the ARIMA model building process.

Autocorrelation

A key concept of ARIMA modeling is autocorrelation. The book *Forecasting Methods and Applications* defines autocorrelation as:

> The correlation between values of the same time series at different time periods. It is similar to correlation, but relates the series for different time lags. Thus there may be an autocorrelation for a time lag of 1, another for a time lag of 2, and so on (Makridakis, Wheelwright, & Hyndman, 1998: 590).

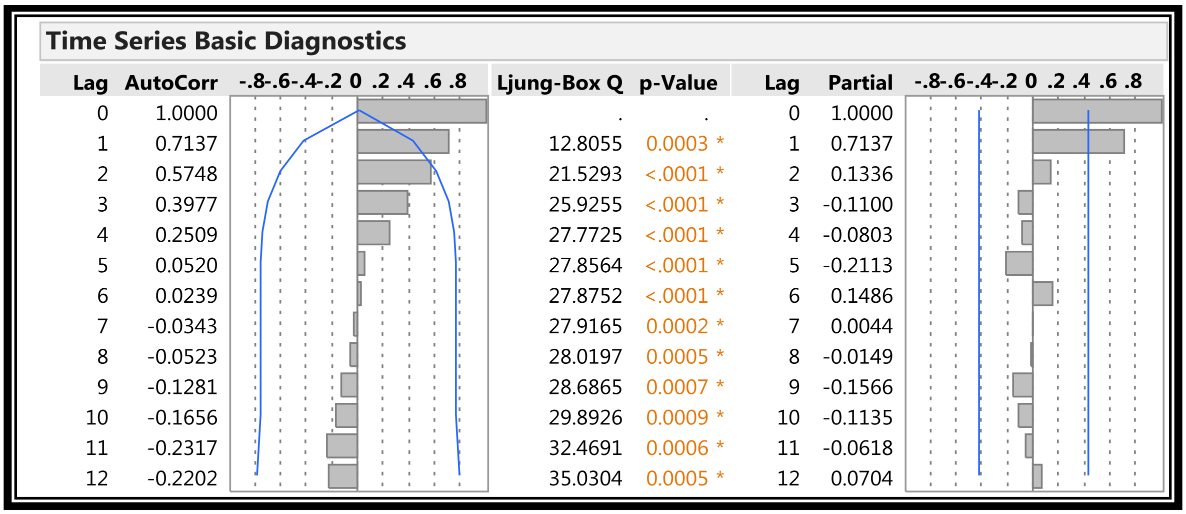
Lag is the separation in time between an observation and a previous observation (Makridakis, Wheelwright, & Hyndman, 1998). Autocorrelation is similar to autoregression, but key differences exist. Autocorrelation is used to assess the relationship of time series data. Whereas autoregression is used to forecast with time series data based on the mathematical relationship autocorrelation describes (Carlberg, 2013). Autoregression is discussed further in the General Non-Seasonal ARIMA Model section.

The key autocorrelation statistic is the autocorrelation coefficient for the *kth* lag (k= the lag number) (Makridakis, Wheelwright, & Hyndman, 1998). The formula is shown in Equation 10; where $\bar{Y}$ is the mean of the number (n) of non-missing points, Y_t is the observation in time (current) while Y_{t-k}, observation at a previous time (lagged by k periods) (Makridakis, Wheelwright, & Hyndman, 1998).

Equation 10: Autocorrelation Coefficient

$$r_k = \frac{1}{n}\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})$$

The autocorrelation function (ACF) contains the autocorrelation coefficients and depicts the pattern of autocorrelation (Carlberg, 2013). The ACF plotted against the lag is called a correlogram and is depicted in Figure 5. In Figure 5, the AutoCorr parameter is the autocorrelation coefficient while the bars graphically depict the autocorrelations.


Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.7137	12.8055	0.0003 *	1	0.7137
2	0.5748	21.5293	<.0001 *	2	0.1336
3	0.3977	25.9255	<.0001 *	3	-0.1100
4	0.2509	27.7725	<.0001 *	4	-0.0803
5	0.0520	27.8564	<.0001 *	5	-0.2113
6	0.0239	27.8752	<.0001 *	6	0.1486
7	-0.0343	27.9165	0.0002 *	7	0.0044
8	-0.0523	28.0197	0.0005 *	8	-0.0149
9	-0.1281	28.6865	0.0007 *	9	-0.1566
10	-0.1656	29.8926	0.0009 *	10	-0.1135
11	-0.2317	32.4691	0.0006 *	11	-0.0618
12	-0.2202	35.0304	0.0005 *	12	0.0704

Figure 5: ACF and PACF Plot

According to the *JMP® 11 Specialized Models* guidebook "the [solid blue] curves show twice the large-lag standard error (+/-) 2 standard errors" for 95% confidence limits (JMP, 2014: 158). A large autocorrelation from a previous lag (k-1) may inflate subsequent lags before dampening (dying out) (Box, Jenkins, & Reinsel, 2008). Because of this phenomenon, an adjustment is made to determine the significant autocorrelation from the inflated value; the large-lag is the adjustment for this interdependence (Box, Jenkins, & Reinsel, 2008). The autocorrelation coefficient standard error (SE_k) is computed with Equation 11, while the large lag standard is the square root of SE_k (Box, Jenkins, & Reinsel, 2008).

Equation 11: Autocorrelation Standard Error

$$SE_{k=} \frac{1}{n}(1 + 2\sum_{i=1}^{k-1} r_i^2)$$

Partial Autocorrelation

The book *Forecasting Methods and Applications* states, "partial autocorrelations are used to measure the degree of association between observations Y_t and Y_{t-k}, when the effects of other time lags (1, 2, 3, …, k-1) are removed" (Makridakis, Wheelwright, & Hyndman, 1998: 320). Makridakis, Wheelwright, and Hyndman further explain, "the partial autocorrelation coefficient of order *k* is denoted by *αk* and can be calculated by regressing *Yt* against *Yt-1*, …, *Yt-k*" (1998: 321). The partial autocorrelation coefficient formula is shown in Equation 12 where the *αk* is represented by the coefficient β_k.

Equation 12: Partial Autocorrelation Coefficient

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_k Y_{t-k}$$

The solid blue lines represent 2 standard errors for 95% confidence limits in the PACF plot (see right side of Figure 5 for an example) (JMP, 2013). The partial autocorrelation coefficient standard error is computed as follows (Makridakis, Wheelwright, & Hyndman, 1998):

Equation 13: Partial Autocorrelation Standard Error

$$SE_k = \frac{1}{\sqrt{n}}$$

White Noise Model

An assumption of ARIMA models is the forecast residuals follow a white noise model (Box, Jenkins, & Reinsel, 2008). According to the book *Forecasting Methods and Applications,* a white noise model "is a simple random model where observation Y_t is made up of two parts, an overall level, c, and a random error term, e_t which is uncorrelated from period to period" (Makridakis, Wheelwright, & Hyndman, 1998: 317). Equation 14 shows the white noise model:

Equation 14: White Noise Model

$$Y_t = c + e_t$$

The white noise model is a critical aspect of time series analysis. In theory, all autocorrelation coefficients of white noise data have a sampling distribution approximately normal with a mean of zero and standard error of $1/\sqrt{n}$, where n is the number of observations (Makridakis, Wheelwright, & Hyndman, 1998). Each lag's mean can be compared to zero with a t-test. Once again, the solid blue lines on the ACF side in Figure 5 represent two standard errors (JMP®, 2013). Values within the blue lines are not statistically different than zero (JMP®, 2013). Values outside the blue lines are statistically different than zero thus we can infer those observations are not random (white noise), they represent a pattern (Box, Jenkins, & Reinsel, 2008). In addition to the white noise model, the sampling distribution is another foundational concept in time series analysis (Makridakis, Wheelwright, & Hyndman, 1998). The distribution and standard error provide insight into what is random (white noise) and what is a true pattern or significant relationship (Makridakis, Wheelwright, & Hyndman, 1998).

Portmanteau Tests

Portmanteau tests allow multiple autocorrelation coefficients to be tested at once (Makridakis, Wheelwright, & Hyndman, 1998). The most common portmanteau tests are the Box-Pierce and Ljung–Box test (Makridakis, Wheelwright, & Hyndman, 1998). Both methods use the following hypothesis test:

- **H_0:** The data are independently distributed. The correlations in the population from which the sample is taken are zero, so that any observed correlations in the data result from randomness of the sampling process.
- **H_a:** The data are not independently distributed, the correlations are significantly different than zero (Box, Jenkins, & Reinsel, 2008).

The test statistic for Box-Pierce is displayed in Equation 15 (Box, Jenkins, & Reinsel, 2008).

Equation 15: Box-Pierce Test Statistic

$$Q = n \sum_{k=1}^{h} r_k^2$$

Where *n* is the number of observations and *h* is the maximum lag considered (Makridakis, Wheelwright, & Hyndman, 1998). Equation 16 displays the formula for the Ljung-Box test statistic (Q^*) which is similar, but slightly different than the Box-Pierce test (Box, Jenkins, & Reinsel, 2008):

Equation 16: Ljung-Box Test Statistic

$$Q^* = n(n+2) \sum_{k=1}^{h} (n-k)^{-1} r_k^2$$

The r_k variable is the autocorrelation value for observation k (Box, Jenkins, & Reinsel, 2008). Both portmanteau tests compare the test statistic (Q and Q*) to the chi-square distribution (χ^2) to determine if the plot of the residuals is statistically different from zero (white noise) or "to test that the residuals from a model can be distinguished from white noise" (JMP, 2013: 158). The Ljung-Box Q^* and p-values appear for each autocorrelation lag as depicted in Figure 5 (JMP®, 2013). A small p-value means the data are significantly different than zero (not random/white noise). We rely on Ljung-Box in this research because the software (JMP®11.0) provides the Ljung-Box (Q*) and theory indicates it has advantages over the Box-Pierce test (Q) (Bowerman & O'Connell, 1993: 497).

Time Series Patterns

There are four patterns in which time series data are categorized: horizontal (stationary), seasonal, cyclical, and trend (Makridakis, Wheelwright, & Hyndman, 1998). A stationary pattern occurs when the observations fluctuate around a constant mean; an example is a product with sales that do not fluctuate much over time (Makridakis, Wheelwright, & Hyndman, 1998). A seasonal pattern exists when certain factors influence the time series; for example, Christmas and other holidays affect the sales of many products. A cyclical pattern exists when the increases and decreases of the data are not due to a fixed period; the lack of a fixed period is what differentiates cyclical from seasonal; examples include industries correlated with the macro-economy and business cycle (steel, automobiles, and major appliances) (Makridakis, Wheelwright, & Hyndman, 1998). A trend pattern exists where there is a long term rise or decline in the data;

examples included sales from many companies, the gross national product, and energy usage (Makridakis, Wheelwright, & Hyndman, 1998). Many data series are comprised of multiple patterns (Makridakis, Wheelwright, & Hyndman, 1998). Given the nature of this research we do not expect to identify any seasonal or cyclical patterns. Although trend patterns may exist we expect to primarily deal with stationary EV indices (CPI, SPI, SPI(t), and BEI).

Examining Stationarity

In time series analysis, stationary essentially means no growth in the data with a constant mean and variance that is independent of time (Makridakis, Wheelwright, & Hyndman, 1998). There are multiple ways to check stationarity. The most basic check is a visual examination of the time series plot. A stationary plot is free of upward or downward trends, with the spikes close to equal distance from the mean so they effectively cancel each other out. Figure 6 graphically depicts a stationary time series.

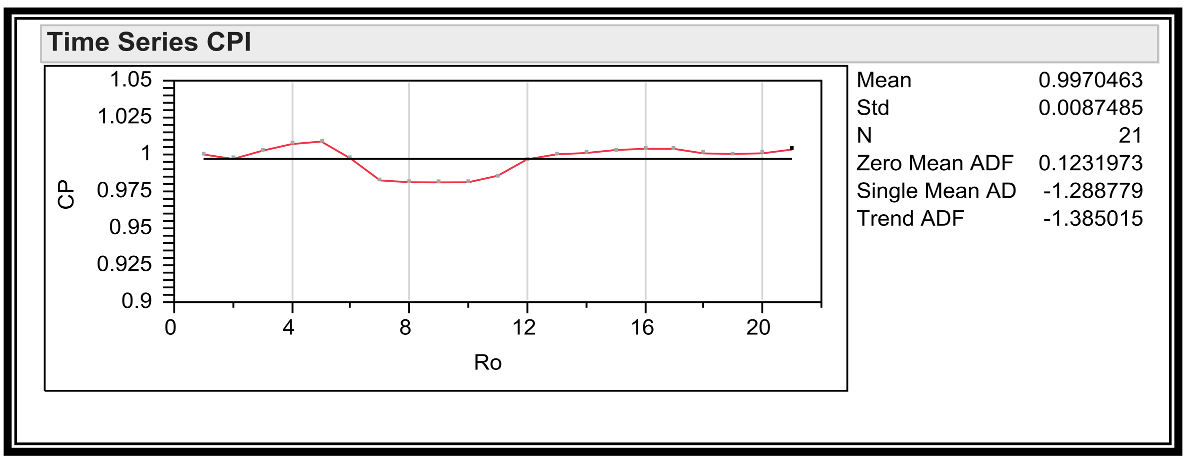


Figure 6: CPI Time Series Graph

Another method to detect stationarity involves examining the ACF plot (Figure 5). According to the book *Forecasting Methods and Applications*, "the autocorrelations of stationary data drop to zero quickly while the non-stationary series will remain significantly different than zero for several time lags" (Makridakis, Wheelwright, & Hyndman, 1998: 326-327). When a visual examination of the ACF plot does not provide conclusive results, the Augmented Dickey-Fuller test (ADF) can be used (JMP®, 2013). The ADF test determines stationarity with a mathematical test statistic and the following hypothesis test (JMP®, 2013):

- **H_0**: Test Statistic = 0 (not stationary)
- **H_a**: Test Statistic < 0 (the data is stationary)

A negative value denotes a stationary time series (JMP®, 2013). The JMP® 11.0 output produces three ADF tests: zero mean, single mean, and trend (2013). Because the indices in this research should never be zero the means will be single or trend. Figure 6 shows negative single and trend ADF test statistics therefore this time series is considered stationary.

Removing Stationarity

When trends or other non-stationary patterns exist in the times series, the resulting positive autocorrelations dominate the ACF plot (Makridakis, Wheelwright, & Hyndman, 1998). Therefore it is critical to remove the non-stationarity in order to assess the true autocorrelation structure before proceeding with the model building process (Makridakis, Wheelwright, & Hyndman, 1998). One approach is called differencing and is defined by the book *Forecasting Methods and Applications* as "the change between each observation in the orignial series. The differenced series will have only n-1 values since it is not possible to calculate a difference (Y'_1) for the first observation" (Makridakis, Wheelwright, & Hyndman, 1998: 326) The differencing calcluation is shown in Equation 17 (Makridakis, Wheelwright, & Hyndman, 1998):

Equation 17: First Order Differencing

$$Y'_t = Y_t - Y_{t-1}$$

Taking the first difference is a useful method for eliminating stationarity (Makridakis, Wheelwright, & Hyndman, 1998). However, the first difference may not remove the stationarity completely. In this case, the data can be differenced again. This series will have n-2 values and contain two integrated (I) parameters. The formula is shown in Equation 18 (Makridakis, Wheelwright, & Hyndman, 1998):

Equation 18: Second-Order Differencing

$$Y"_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

General Non-Seasonal ARIMA Model

According to the book *Predictive Analytics*, the term ARIMA stands for:

- AR: Autoregressive. The model and forecast can be partially or completely based on autoregression.
- I: Integrated. The baseline may need to be differenced and the differenced series modeled. In order to forecast, the difference(s) are reversed by a process called integrating. This restores the baseline to its original level.
- MA: Moving Average. Not based on an average of observations, but an average of a model's errors (Carlberg, 2013: 242)

Regression with time lagged input variables is called autoregression (AR) and is based on the general form of Equation 19 (Makridakis, Wheelwright, & Hyndman, 1998).

Equation 19: Autoregression

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots \beta_p Y_{t-p} + e_t$$

Conceptually, AR is similar to regression; the difference is the response variables from previous periods are used as explanatory variables to compute the current period's response (Yt) (Makridakis, Wheelwright, & Hyndman, 1998).

As previously discussed, the residuals (or error terms) can also be used as explanatory variables in a regression equation (Makridakis, Wheelwright, & Hyndman, 1998):

Equation 20: Moving Average (Box-Jenkins)

$$Y_t = \beta_0 + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \cdots \beta_q e_{t-q} + e_t$$

Here the dependence relationship among successive error terms ($e_{t-1}, e_{t-2}, \ldots e_{t-q}$) is called a moving average (MA) model (Makridakis, Wheelwright, & Hyndman, 1998). This is obviously different than a simple moving average which is an average of observed values. To avoid confusion, this research only uses the term moving average (MA) when referring to ARIMA models.

Autoregressive (AR) and moving average (MA) parameters can be combined to form autoregressive moving average (ARMA) models (Makridakis, Wheelwright, & Hyndman, 1998). ARMA models can only be used with stationary data; if the original data is non-stationary, the data must be differenced (Makridakis, Wheelwright, & Hyndman, 1998). At this point, the model is now called an autoregressive integrated moving average (ARIMA) model. There are a large number of possible ARIMA models. The general non-seasonal model is known as ARIMA (p, d, q) (Carlberg, 2013):

- AR: p = number of the autoregressive parameters in the model

- I: d = the number of times the data has been differenced to achieve stationarity
- MA: q= the number of moving average parameters in the model (Carlberg, 2013: 243)

A white noise model is classified as ARIMA (0,0,0); while a random walk model is classified as ARIMA (0,1,0) or I(1) because it has one degree of differencing and no AR or MA parts (Makridakis, Wheelwright, & Hyndman, 1998).

The simplest AR model is the first order ARIMA (1,0,0) which is also denoted by AR(1). The equation is mathematically defined in Equation 21 where observation *Yt* depends on *Yt-1* with the coefficient $\emptyset_1$ restricted to -1 to 1 (Makridakis, Wheelwright, & Hyndman, 1998: 337). The time series is equivalent to a white noise model when $\emptyset_1$ = 0. When $\emptyset_1$= 1, the time series is equivalent to a random walk model (Makridakis, Wheelwright, & Hyndman, 1998: 337-338).

Equation 21: ARIMA (1,0,0)

$$Y_t = c + \emptyset_1 Y_{t-1} + e_t$$

The simplest MA model is the first order ARIMA(0,0,1) or MA(1). The model is mathematically defined in Equation 22 where observation Y_t depends on the residual (e_t) and also the previous residual (e_{t-1}); the coefficient is restricted to lie between -1 and 1 (Makridakis, Wheelwright, & Hyndman, 1998).

Equation 22: ARIMA (0,0,1)

$$Y_t = c + e_t - \theta_1 e_{t-1}$$

In practice it is rarely necessary to use values other than 0, 1, or 2, because this small range of values covers a great range of forecasting situations (Makridakis, Wheelwright,

& Hyndman, 1998). Now that the essential concepts have been discussed we can move to the model building process itself.

Box-Jenkins Approach

This section will describe the three phases of the Box-Jenkins methodology: Identification, Estimation and Testing, and Application. Figure 7 visually depicts the Box-Jenkins methodology (Makridakis, Wheelwright, & Hyndman, 1998).

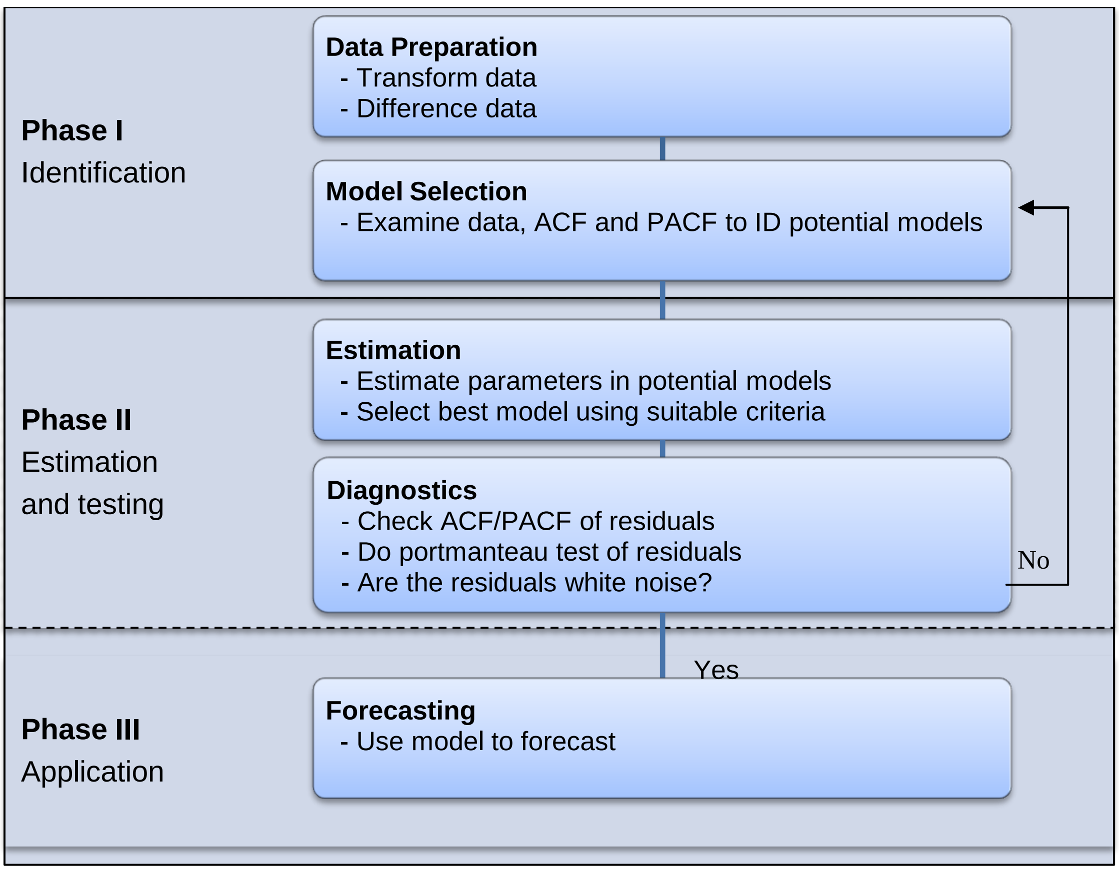


Figure 7: Box-Jenkins Methodology Flowchart

Phase I – Identification

As the name implies the objective of this phase is to identify models that are potentially suitable for the time series data being analyzed. Data preparation and model selection takes place in this phase. Makridakis, Wheelwright, and Hyndman recommend the following steps for phase one (1998: 347):

1. Plot the time series data
2. Assess the data for stationarity
3. Use differencing if the series is not stationary
4. Once stationarity is achieved, examine the ACFs and PACFs to assess patterns with three possibilities to consider.
    a. Does seasonality exist
    b. AR or MA model may be determined
    c. If AR or MA is not clearly suggested, an ARIMA may be necessary

The first three steps have been discussed in the previous sections. Seasonality is not a concern (4.a.), but steps 4.b. and 4.c. are crucial in the identification phase. To identify a suitable model we compare the observed patterns with the theoretical (expected) ACF and PACF patterns with the approach outlined in Table 5 (Makridakis, Wheelwright, & Hyndman, 1998; Montgomery, Johnson, & Gardiner, 1990). Within Table 5 the expression *tails off* means the function (ACF, PACF) decays in an exponential, sinusoidal (sine wave), or geometric fashion with potentially more nonzero values than zero (Montgomery, Johnson, & Gardiner, 1990). Whereas *cuts off* refers to the function truncating abruptly to zero with few nonzero values (Montgomery, Johnson, & Gardiner, 1990). In the previous sentences, zero denotes within (+/-) 2 standard errors (not statistically different than zero). A nonzero value is outside the (+/-) 2 standard errors (statistically different than zero). Table 5 highlights the dichotomy between AR and MA models. In an AR the ACF tails off while the PACF cuts off. In an MA the ACF cuts off while the PACF tails off. With this in mind the combined ARMA model contains a tail off for both ACF and PACF.

An ARIMA (p,d,q) model is an option if no clear AR, MA, or ARMA model is delineated. The general ARIMA models yields a great variety of patterns in the ACF and PACF; given this fact, there are no clear rules for visually identifying ARIMA models (Makridakis, Wheelwright, & Hyndman, 1998). If differencing is required (non-stationary data) an ARIMA model is a logical choice, otherwise choosing the specific model type (p,d,q) is based on judgment, experience, and experimentation (trial and error).

Table 5: Expected Patterns in the ACF and PACF for AR and MA Models

Process	**ACF**	**PACF**
AR (1)	*Tails off* (Exponential decay): • positive if $\varphi_1 > 0$ • alternating in sign starts (-) if $\varphi_1 < 0$	*Cut off* (spike at lag 1, then cuts to zero) • spike is positive if $\varphi_1 > 0$ • spike is negative if $\varphi_1 < 0$
AR(p)	*Tails off* (Exponential decay or damped sinewave)	*Cuts off* after lag p
MA (1)	*Cuts off* (Spike at lag 1, then cuts to zero): • spike is positive if $\theta_1 < 0$ • spike is negative if $\theta_1 > 0$	*Tails off* (Exponential decay): • negative if $\theta_1 > 0$ • alternating in sign starts (+) if $\theta_1 < 0$
MA(q)	*Cuts off* (spikes at lags 1 to q then cuts off after lag q)	*Tails off* (Exponential decay or damped sinewave)
ARMA(p, q)	*Tails off* (Exponential decay)	*Tails off* (Exponential decay)

The potential models are identified by first setting boundaries on the ARIMA parameters. As previously discussed, it is generally not necessary to use parameters greater than two. Restricting the ARIMA parameters to the values listed in Table 6 yields the 27 models listed in Table 7. The next phase of the Box-Jenkins methodology is Phase II (estimation and testing).

Table 6: ARIMA Model Parameters

ARIMA parameter	Minimum	Maximum
p	0	2
d	0	2
q	0	2

Table 7: Potential ARIMA Models

AR(1)	AR(2)	MA(1)	MA(2)
ARMA(1, 1)	ARMA(1, 2)	ARMA(2, 1)	ARMA(2, 2)
ARI(1, 1)	ARI(1, 2)	ARI(2, 1)	ARI(2, 2)
IMA(1, 1)	IMA(1, 2)	IMA(2, 1)	IMA(2, 2)
ARIMA(0, 0, 0)	ARIMA(1, 1, 1)	ARIMA(1, 1, 2)	ARIMA(1, 2, 1)
ARIMA(1, 2, 2)	ARIMA(2, 1, 1)	ARIMA(2, 1, 2)	ARIMA(2, 2, 1)
ARIMA(2, 2, 2)	I(1)	I(2)	

Phase II – Estimation and testing

In this phase the parameters are estimated in potential models, then the best model is selected based on suitable criteria. Finally, diagnostic tests are conducted to ensure the model meets the underlying assumptions. With our list of potential models from Table 7 we can use computer programs to find appropriate initial estimates. The software used in this research is JMP ® version 11. The *JMP Specialized Models* guidebook explains the estimation process, “the [ARIMA] models are fit by maximizing the likelihood function, using a Kalman filter to compute the likelihood function” (JMP®, 2013: 162).

For each parameter estimate ($\hat{\theta}$) there is also a standard error ($s_{\hat{\theta}}$) (Bowerman & O'Connell, 1993). A significance test is conducted with these two values with (alpha =

0.05). The t-ratio is shown in Equation 23 utilizing the following hypothesis test (Bowerman & O'Connell, 1993):

- **H_0**: $\theta = 0$. The parameter is equal to zero (not significantly different than zero).
- **H_a**: $\theta \neq 0$. The parameter is not equal to zero (significantly different than zero).

Equation 23: ARIMA Parameter Test Statistic

$$t = \frac{\widehat{\theta}}{s_{\widehat{\theta}}}$$

If the p-value is less than alpha, the parameter is not equal to zero (significantly different than zero). If the p-value is greater than alpha the parameter is not significantly different than zero. Generally, a t-ratio of at least 2 in absolute value will be considered significant (JMP, 2013: 166). The AR parameter was tested for significance as exhibited in Figure 8. In this example, the parameter is not equal to zero (0.0001 < 0.05), therefore this model's AR (1) parameter is significant.

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	0.81577641	0.0650184	12.55	<.0001 *	0.17764315
Intercep	0	0.96428017	0.0233158	41.36	<.0001 *	

Figure 8: AR Model Parameter Estimates

Model Rank

There may be more than one valid model out of the twenty-seven considered. We need a method to determine the best model. The recommended approach is a method that prevents over-fitting by adding a penalty for adding more explanatory variables. For ARIMA models the likelihood (L) is penalized for added terms (parameters) (Makridakis, Wheelwright, & Hyndman, 1998). Two criteria are provided by JMP ® 11: the Akaike's Information Criterion (AIC) and the Schwarz's Bayesian Criterion (SBC or BIC) (2013). These measures are computed as follows (Makridakis, Wheelwright, & Hyndman, 1998):

Equation 24: Akaike's Information Criterion (AIC)

$$\mathrm{AIC} = -2\log\mathrm{L} + 2\mathrm{m}$$

Equation 25: Schwarz's Bayesian Criterion (SBC)

$$\mathrm{SBC} = -2\log\mathrm{L} + \mathrm{m}\ln(\mathrm{n})$$

Where *n* is the number of observations and *m* = the number of parameters in the model (including the intercept) (Makridakis, Wheelwright, & Hyndman, 1998). Lower AIC or SBC values indicate a better fitting model (JMP, 2013). Figure 9 depicts an individual model summary whereas Table 8 summarizes multiple models. Out of the eight models compared, the AR(1) has the lowest AIC and SBC. Therefore AR(1) is deemed the best model. The AIC and SBC are similar measures, for simplicity this research uses the lowest AIC to select the best model.

Model Summary

DF	69	Stable	Ye
Sum of Squared Errors	0.10454555	Invertibl	Ye
Variance Estimate	0.00151515		
Standard Deviation	0.03892497		
Akaike's 'A' Information Criterio	-256.39331		
Schwarz's Bayesian Criterion	-251.86795		
RSquare	0.68090163		
RSquare Adj	0.67627702		
MAPE	2.1452803		
MAE	0.01958656		
-2LogLikelihood	-260.39331		

Figure 9: AR Model Summary

Table 8: ARIMA Model Comparison

Model	DF	Variance	AIC	SBC	R Square	-2LogLH	Weights	MAPE	MAE
AR(1)	69	0.00151	-256.39	-251.86	0.681	-260.39	0.6125	2.145	0.019
ARMA(1, 1)	68	0.00153	-254.69	-247.90	0.682	-260.69	0.2616	2.178	0.019
ARIMA(1, 1, 1)	67	0.00147	-252.48	-245.73	0.676	-258.48	0.0866	2.312	0.021
IMA(1, 1)	68	0.00162	-248.83	-244.33	0.658	-252.83	0.0139	2.267	0.021
ARI(1, 1)	68	0.0016	-248.68	-244.18	0.657	-252.68	0.0129	2.236	0.020
I(1)	69	0.00165	-248.57	-246.32	0.651	-250.57	0.0122	2.235	0.020
MA(1)	69	0.00254	-220.27	-215.74	0.452	-224.25	0	3.480	0.031
ARIMA(0, 0, 0)	70	0.00468	-178.35	-176.095	0	-180.35	0	4.986	0.045

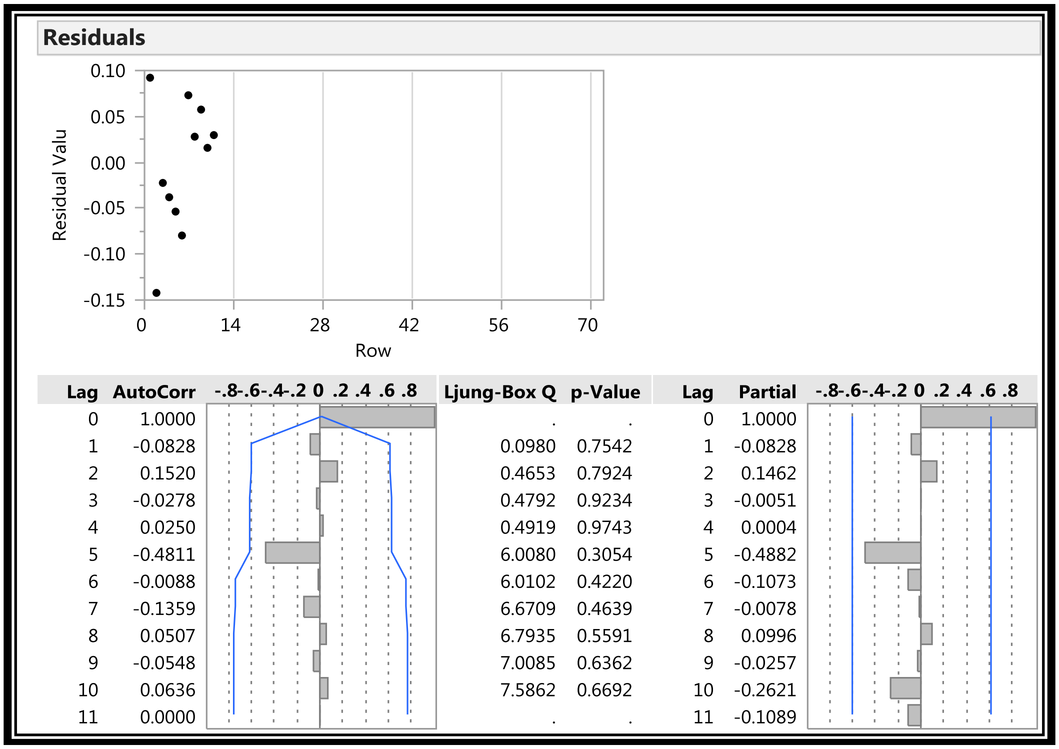
Diagnostic Checking

Now that we have chosen the best model, the following diagnostics must be conducted to determine if the residuals are white noise (Makridakis, Wheelwright, & Hyndman, 1998). The objective is to find no significant autocorrelations or partial autocorrelations when checking the residuals' ACF and PACF (Makridakis, Wheelwright, & Hyndman, 1998). The first step is a visual inspection of the residuals' ACFs and PACFs plot. If any ACFs or PACFs (except lag 0) are outside the acceptable range we reject the null and conclude the model's residuals are not white noise. The next step is an additional check that involves the Ljung-Box test of the following hypothesis:

- **H_0:** The residuals are independently distributed. The residuals are white noise.
- **H_a:** The residuals are not independently distributed; the residuals are not white noise.

If the p-value is less than alpha (0.05) we reject the null, if it is greater than alpha we fail to reject the null and conclude the residuals are white noise. In the example from

Figure 10, the p-values are greater than the alpha. We fail to reject the null and conclude that the residuals are from white noise. Once the diagnostic checks are passed the model is deemed adequate, therefore it is not necessary to further modify the model (Makridakis, Wheelwright, & Hyndman, 1998). The model can now be used to forecast.



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.0828	0.0980	0.7542	1	-0.0828
2	0.1520	0.4653	0.7924	2	0.1462
3	-0.0278	0.4792	0.9234	3	-0.0051
4	0.0250	0.4919	0.9743	4	0.0004
5	-0.4811	6.0080	0.3054	5	-0.4882
6	-0.0088	6.0102	0.4220	6	-0.1073
7	-0.1359	6.6709	0.4639	7	-0.0078
8	0.0507	6.7935	0.5591	8	0.0996
9	-0.0548	7.0085	0.6362	9	-0.0257
10	0.0636	7.5862	0.6692	10	-0.2621
11	0.0000	.	.	11	-0.1089

Figure 10: Plots of ACF and PACF for Residuals

Phase III – Application

Forecasting with the model is straight forward. The prediction equation will depend on the model type selected. In practice the user chooses the model based on the previous steps then relies on the software to calculate the forecasted values. The forecast values are based on the number of significant lags and forecasted periods. With the exception of the intercept, the number of lags must be more than one, but less than the number of observations (see Figure 8). In Phase I the user can decide the number of lags to be considered. Unlike in the beginning of the analysis, the software does not allow the user to change the lagged periods used in the prediction formula. This research uses the model's first forecasted value (next month) as a performance factor in the time estimate formula (time estimate = planned duration/performance factor (PF)).

Time Series Summary

In 2011, AFIT student C. Grant Keaton used time series analysis to detect changes in the CPI and SPI to evaluate a contract's performance. This literature review has not discovered any studies that applied time series analysis to forecast the duration of DoD programs. In this research, time series analysis is used to forecast values based on previous period's data rather than the current period's index value (SPI, SPI(t) or CPI). If the pattern from previous periods is different than the cumulative index value then the forecasted value will be different. The difference will lead to different and possibly more accurate duration forecasts.

The Box-Jenkins approach is a robust method and is easy to implement if the user has access to the proper software. The strength is the systematic procedure used to determine the model that best fits the data. Given this robustness, ARIMA models are arguably the most accurate time series forecasting method (Montgomery, Johnson, & Gardiner, 1990). Beyond the assumptions already listed, ARIMA models, like all models, have weaknesses. On the technology side, many practitioners will not have access to JMP® or other powerful statistical software. The open source R statistical software contains the capability to conduct time series analysis, but it may have a steeper learning curve than commercial off the shelf software. The book *Predictive Analytics* by

Carlberg (2013) provides software add-ins that make times series analysis easier in Excel; unfortunately, that package is not as efficient as JMP ®11. In addition to the software concerns, some of the time series concepts are complex thus making this method inaccessible if the practitioner does not have a working knowledge of forecasting. The largest potential downside is complex techniques such as ARIMA models are not guaranteed to significantly improve accuracy over simpler techniques (Makridakis, Wheelwright, & Hyndman, 1998).

Schedule Forecasting: Kalman Filter Forecasting Method

In 2007, Kim developed a new schedule forecasting technique, the Kalman filter forecasting method (KFFM). The KFFM assesses a project's progress and calculates a probability distribution for the duration at completion (Kim, 2007). In simple terms, the KFFM is a hybrid of Earned Schedule (ES) and a Kalman filter (Kim, 2007). According to Kim, "the Kalman filter is a recursive algorithm used to estimate the true state, but hidden state of a dynamic system using noisy observations (2007: 23). Rudolph Kalman wrote the seminal paper in 1960; the Kalman filter has been applied to broad areas including autonomous or assisted navigation (Welch & Bishop, 2001). The Kalman filter application to schedule estimating is relatively new and has not been applied to DoD programs (Kim, 2007). The KFFM provides a probabilistic framework that incorporates actual performance data being generated by a project (earned value) and prior knowledge of the program (planned value) to forecast the project's future progress (Kim & Reinschmidt, 2010).

The Kalman filter approach used by Kim will be discussed (2007). The foundation of the KFFM is a recursive algorithm that uses prior and posterior information to continuously update estimates via a learning cycle shown in Figure 11 (Kim, 2007; Kim & Reinschmidt, 2010). Within the Kalman filter framework, the state of the dynamic system is represented by two sets of variables: the state variables (x_k) and the error covariance variables (P_k) (Kim, 2007; Kim & Reinschmidt, 2010). The error covariance is a measure of the uncertainty in the estimates of the state variables (Kim, 2007). According to Kim, "the states and covariance are updated through two stochastic linear models: the measurement model and the system model" (2007: 24). The measurement model updates the prior estimate with new information (z_k) to correct the estimate (resulting in the posterior estimate) (Kim, 2007). Kim further describes the process as "the system model predicts the future state of the system at the next time period" (2007: 24).

KFFM Process

Figure 11 outlines the KFFM process while Table 9 lists the variables and equations used in Kim's study (2010). The process begins with the initial estimates of the state vector and error covariance (Kim & Reinschmidt, 2010). The state vector is a 2x1 matrix: the time variance at time *k* (TV*k*) and its rate of change from the previous period (dTV*k* / dt) (Kim & Reinschmidt, 2010). The initial state vector (x_k) and error covariance (P_0) are estimated as zero because it is assumed the known uncertainty is incorporated (Equation 26) (Kim & Reinschmidt, 2010).

Equation 26: Kalman Filter Initial Estimates

$$\hat{x}_k = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \quad P_0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

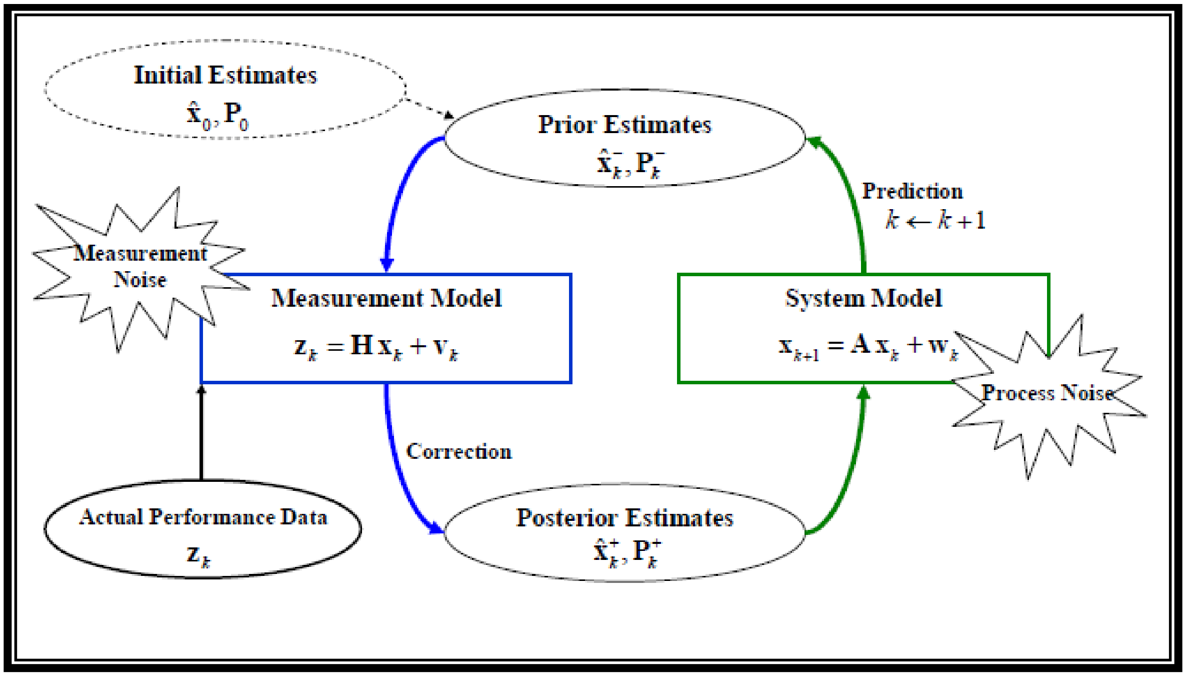


Figure 11: Recursive Learning Cycle of the Kalman Filter

The process noise variable Q adjusts the Kalman gain (K); the Q is estimate based on the mean of the initial estimated duration (Kim & Reinschmidt, 2010). The initial estimate can be derived from a three point Program Evaluation and Review Technique (PERT) estimate, listed in Equation 27 and Equation 28 (Kim & Reinschmidt, 2010). In this example the process noise (q) equals 0.694 (the variance is $(0.83)^2 = 0.694$) (Equation 29).

Equation 27: PERT Estimate (Mean)

$$\text{Mean} = \frac{\text{O} + 4\text{M} + \text{P}}{6} = \frac{0.95 * 50 + 4 * 50 + 1.05 * 50}{6} = 50 \text{ months}$$

Equation 28: PERT Estimate (Standard Deviation)

$$\text{Standard Deviation} = \frac{\text{P} - \text{O}}{6} = \frac{1.05 * 50 - 0.95 * 50}{6} = 0.83 \text{ months}$$

Equation 29: Process Noise Matrix

$$Q_k = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix}$$

Table 9: Kalman Filter Forecasting Model Components

Components	Equations	Description
State vector	$x_k = \begin{Bmatrix} TV_k \\ dTV_k/dt \end{Bmatrix}$	TV*k* = TV that is defined as the earned schedule minus the time of forecasting.
Dynamic system model	$x_k = A_k x_k - 1 + w_{k-1}$ $A_k = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix};\ \ w_{k-1} = \begin{bmatrix} 0 \\ w_{k-1} \end{bmatrix}$	Ak=transition matrix. w*k*−1=vector of random process noise and *wk*−1=random error term for the derivative of the TV.
Measurement model	$z_k = Hx_k + v_k$ $z_k = [z_k]; H = \{1\ 0\}; v_k = [v_k]$	H=observation matrix. v*k*=vector of random measurement noise and *vk*=random error term for the measurement *zk*.
Prediction process	$\hat{x}_k^- = A\hat{x}_{k-1}^+$ $P_k^- = AP_{k-1}^+ A^T + Q_{k-1}$	Before observing a new TV*k* at time period *k*, the prior estimates of the state vector and the error covariance matrix P are calculated. Q*k*−1=process noise covariance matrix.
Kalman gain	$K_k = P_k^- H^T (HP_k^- H^T + R_k)^{-1}$	Kalman gain at time period *k*, which is determined in such a way that minimizes the posterior error covariance matrix. R*k*=measurement error covariance matrix.
Updating process	$\hat{x}_k^+ = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$ $P_k^+ = [I - K_k H]P_k^-$	The posterior estimates of the state vector and the error covariance matrix are calculated using the Kalman gain.

The variance of measurement error is the error associated with the measurement process (v_k); unless known, this variable is also estimated with PERT (Equation 30) (Kim & Reinschmidt, 2010):

Equation 30: Variance of Measurement Error

$$\text{Variance of } v_k = \left[\frac{a-(-a)}{6}\right]^2 = \frac{a^2}{9}$$

$$\text{Maximum } v_k = a$$

$$\text{Minimum } v_k = -a$$

Kim and Reinschmidt used a measurement error of ± 3 months so $R_k = 1.0$ (2010). This value can be increased or decreased based on the program manager's confidence in the reliability of the data source (Kim & Reinschmidt, 2010). The R_k simplifies to the r (the measurement error variable) displayed in Equation 31 (Kim & Reinschmidt, 2010):

Equation 31: Measurement Error Matrix

$$R_k = [r]$$

The steps outlined in Figure 11 and the calculations listed in Table 9 have been programmed into KEVM Lite ©, a Microsoft Excel based tool developed by Kim (2010). KEVM Lite © is used in this research to compute duration estimates.

KFFM Applied to Schedule Forecasting

Kim's study used EVM data as the inputs for the KFFM (2007). Specifically the following parameters are used: budget at completion (BAC), planned value (PV), earned value (EV), planned duration (PD), and the reporting date (t). Then Earned Schedule (ES) is used as an input into the estimated duration at completion EDAC (t) formula. The EDAC (t) is forecasted at a point in time (t), which is each month in this study (Kim & Reinschmidt, 2010). The KFFM applies an algorithm to ES and EVM data to predict three EDAC (t) curves shown in Figure 12: the mean, the upper bound, and the lower bound (Kim & Reinschmidt, 2010). Additionally, a probability of schedule slippage (PrSS) is computed. In the same 2010 study Kim and Reinschmidt used two real projects (a gas plant and a refinery plant) to show the KFFM in action.

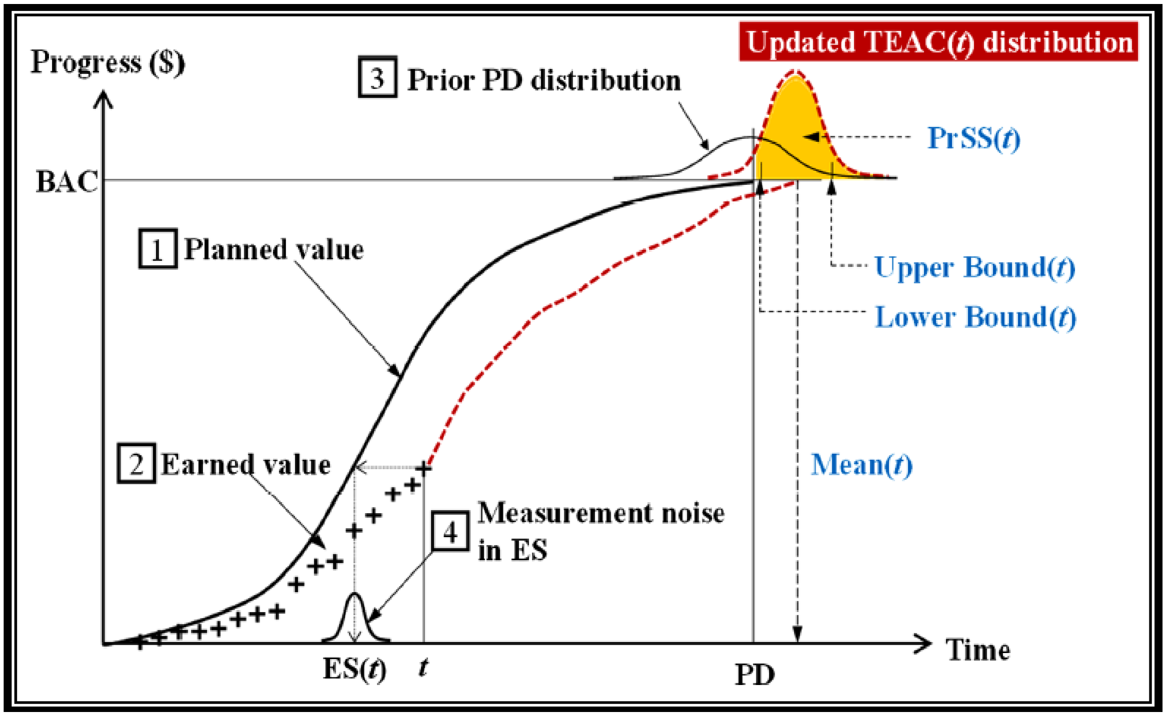


Figure 12: Kalman Filter Forecasting

Kim and Reinschmidt compared the Earned Schedule (ES) method (PD/SPI(t)) to the KFFM (2010). In that study, the KFFM outperforms the ES method in terms of consistent estimates; furthermore, the ES method shows erratic tendencies in the monthly trend analysis (Kim & Reinschmidt, 2010). Kim and Reinschmidt state, "improved forecasting methods based on proven state-of-the art techniques should lead to better project management decisions and improved project performance" (2010: 842). Although the study has merit, it is not without limitations. The primary limitation is a small sample size (two projects). For the purposes of this thesis, another limitation of the Kim and Reinschmidt study is the relatively short planned durations of the projects studied (24 and 25 months). Additionally, DoD programs were not examined. This research will apply the KFFM to lengthier projects and a different type of project (DoD).

Schedule Forecasting: Improving the Planned Duration Estimate

According to the *GAO Schedule Assessment Guide*, "the baseline schedule includes… original forecasts for activity start and finish dates, … original estimates for work, resource assignments, critical paths, and total float [slack] (2012: 136). The current schedule includes new tasks (added since the baseline schedule) and should include updates from actual performance data to forecast the remaining work (GAO, 2012).   Using the baseline schedule as a benchmark to assess the project's schedule performance is a GAO best practice (GAO, 2012). Lastly, the baseline schedule is used with the critical path method (CPM) to estimate the project's duration (Integrated Master Schedule (IMS) planned duration).

In 2014, Lofgren introduced an approach to improve the IMS planned duration estimate. Lofgren argues the importance of the baseline schedule plan on three points: the planners know the major activities, well defined process exists to develop the system, and the Integrated Baseline Review (IBR) allows the contractor and program office to agree on the reasonableness of the baseline plan (2014: 3). Therefore a project's baseline from the initial IMS is an important benchmark for the entire project. Lofgren analyzed 12 MDAP contracts with 133 schedule observations (individual IMSs) (2014: 2). Supporting chapter one's discussion on schedule growth, Lofgren found many schedule estimates were overly optimistic compared to actual performance. In this study, schedule performance (completing tasks on time) rarely improves with project maturity (Lofgren, 2014). Aside from schedule performance, the overall health of the schedule can provide insight. One health check is the % of tasks are coded as hard constraints with a goal of less than 5%. Lofgren's study discovered the majority of IMSs did not meet the hard constraint metric (2014: 6). Another health check is schedule logic; every task must have a predecessor or successor (GAO, 2012). The metric is met if the project has less than 5% of its tasks with missing predecessors and successors (Lofgren, 2014: 7). An IMS that does not meet the 5% metric indicates an improperly maintained plan and is likely to lead to inaccurate duration estimates. In spite of the relatively poor quality of the IMSs, Lofgren not only attempts to improve the accuracy of the estimated completion date (ECD), he also attempts to provide the ECD earlier in the project (2014: 7).

Lofgren's framework relies on a proposed metric, schedule slip, which is added to the planned duration estimate. The first step of this process sets the baseline as the benchmark (Lofgren, 2014). Each subsequent month's IMS data was compared to the baseline IMS to determine the schedule slip; the schedule slip is added to the reported completion date as depicted in Figure 13 (Lofgren, 2014).

The schedule slip metric displayed in Table 10 was derived from Lofgren's framework (2014). In this example, 4.2 months are added to the IMS planned duration of 49.1 months for a total of 53.3 months. For comparison purposes, the contractor performance (CPR) planned duration value was 49.0 months. The following is a list of equations used to develop Table 10.

Equation 32: Schedule Slip

$$\text{Schedule Slip} = \text{Max}\ [\text{Current Finish Date} - \text{Baseline Finish Date} - \text{Total Slack}]$$

Equation 33: IMS Planned Duration

IMS planned duration = start date from CPR to IMS reported end date

Equation 34: CPR Planned Duration

CPR Planned Duration = start date from CPR to Estimated Completion Date from CPR

Equation 35: Independent Duration Estimate

Independent Duration Estimate = IMS planned duration + schedule slip estimate

Equation 36: Enhanced IDE

Enhanced IDE = IDE/PF

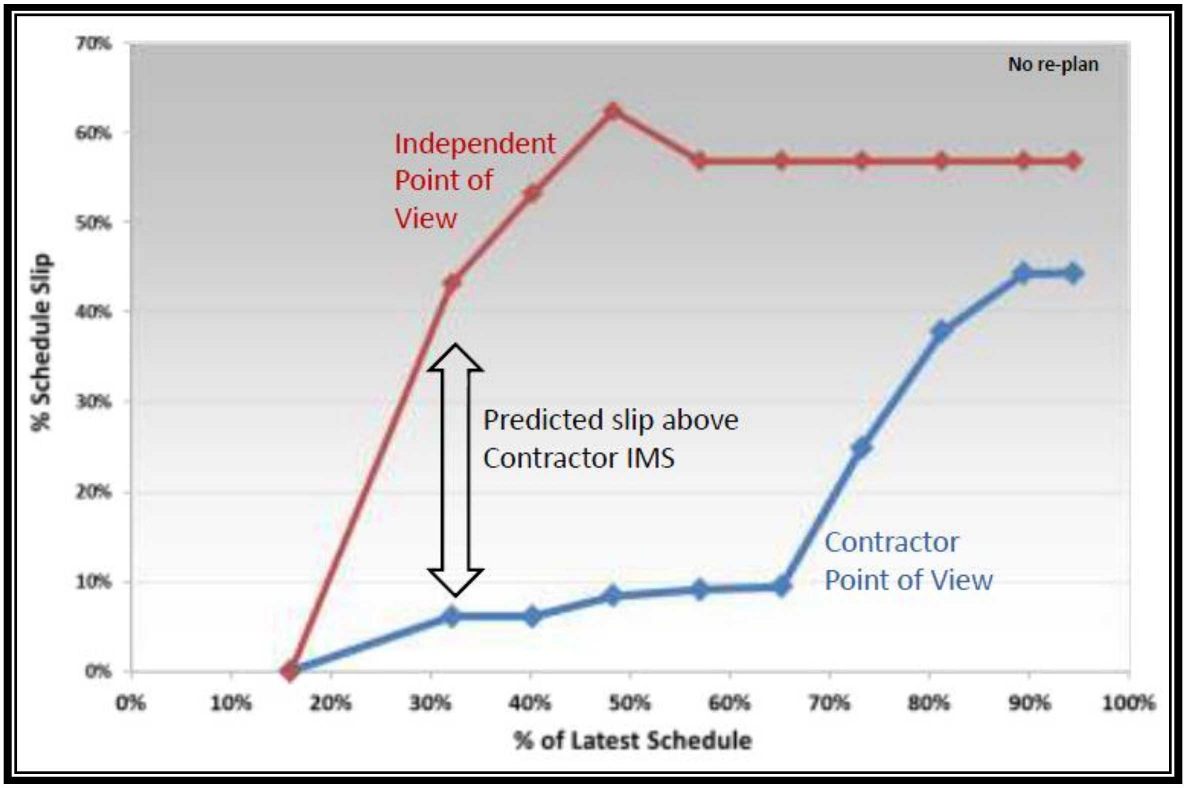


Figure 13: Schedule Slip Method

Incorporating the IMS PD and the baseline analysis by Lofgren appears to be an improvement over the IMS PD by itself. Lofgren's study demonstrated improved accuracy and timeliness over the contractor's reported duration estimate. Although the commodity and contract type were not mentioned, the database is comprised of MDAP contracts thus the results may be generalizable to this research. A weakness of this approach is it is labor intensive. The key argument against conducting an in-depth schedule analysis is that is a labor intensive process. If time is scarce it may make more sense to only use this approach when the IMS PD changes. This approach may reduce the task frequency from monthly to quarterly. Another potential weakness is the fact that the baseline IMS is not usually available until after the integrated baseline review (IBR) (3 to 6 months into the contract) which may make this technique less useful for short duration contracts.

Table 10: IMS Analysis (Current Month Compared to Baseline)

Task Name	Baseline Finish (IMS #1) [4/15/08]	Baseline Total Slack	Current Finish (IMS #2) [5/20/08]	Finish Variance (days)	Slip (days)	Slip (months)
ASIC Build 1-2-3-4 Integration	01/30/08	9	05/02/08	92	83	2.77
PSP Develop Test Cases 1	06/02/08	-47	06/02/08	0	47	1.57
IO : Det Design (S2) Ph 1	05/16/08	-80	07/02/08	46	126	4.20
				MAX	126	4.20

This research uses Lofgren's framework; the schedule slip is added to the current IMS planned duration to obtain an independent duration estimated (IDE), then the IDE and the performance factors are used to calculate an enhanced IDE (Enhanced IDE = IDE/PF).

Baseline Execution Index (BEI)

Related to Lofgren's method is the concept of the Baseline Execution Index (BEI). The Baseline Execution Index (BEI) is a trend metric defined as "the ratio of [baseline] activities that were completed to the number of [baseline] activities that should have been completed by the status date" (GAO, 2012: 148). Three outcomes can be concluded based on the value of the BEI (GAO, 2012: 148):

- BEI = 1 (the project is adhering to schedule)
- BEI < 1 (the project is behind schedule)
- BEI > 1 (the project is ahead of schedule)

The BEI does not measure a project's overall task completion per se, it is concerned with the completion of only the baselines tasks. Eventually as the project matures the BEI will converge to one possibly reducing the metric's usefulness in the late stages of a contract. This phenomenon is a weakness comparable to the SPI. The BEI relies on the concept that the baseline plan is important to the overall performance of the project. With that in mind, the BEI is used as a performance factor (PF) in this research. The BEI was calculated with the National Aeronautics and Space Administration's (NASA) Schedule Test and Assessment Tool (STAT) and the IMS. STAT is a Microsoft ® Project add-in. Finally, the BEI is considered an EVM metric. However, the BEI was not discussed in the forecasting literature. This research attempts to fill the void in the literature.

Summary

In this chapter the relevant literature was reviewed to determine the existing methods used to forecast project duration. Based on this research, Earned Schedule appear to be the best EV index based method. Although ES has been studied extensively, its use in forecasting DoD program duration has not been studied as frequently. The application of time series analysis with EVM data has been studied on a limited basis in the DoD (Keaton, 2011). However, using time series analysis and EVM data to forecast the duration of space programs has not been studied. The KFFM has been used successfully for a limited number of construction type projects, but not for DoD projects. IMS analysis is a recent addition to developing duration estimates; further research is necessary to validate the method on space and development contracts. Finally, using the BEI to forecast duration does not appear in the literature. This research will attempt to fill these voids in the literature by using EVM index based methods (CPI, SPI, SPI(t), and BEI), time series forecast based on EVM indices (CPI, SPI, SPI(t), and BEI), Kalman filter forecasts based on Earned Schedule, and IMS analysis to develop independent duration estimates (IDEs). In the next chapter the specific methodology for each technique is discussed.

III. Methodology

Chapter Overview

This analysis uses Contractor Performance Report (CPR) data to develop schedule estimating models. The purpose of this chapter is to discuss the approaches used to develop the estimating models. First, the data, data source, and data limitations are discussed. Next, the forecasting methods are described: EVM index based, EVM index based plus time series, regression, Kalman filter, and the independent duration estimate (IDE). Finally, the evaluation section explains how the duration forecasting models are evaluated.

Data and Data Source

The EVM Central Repository (EVM-CR) is the primary source of data for this research. The Defense Cost and Resource Center (DCARC) website describes the EVM-CR as a joint effort between DCARC and Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD/AT&L), and is managed by Performance Assessment and Root Cause Analysis (PARCA) (Defense Cost and Resource Center (DCARC), 2014). The EVM-CR provides:

- Centralized reporting, collection, and distribution for key acquisition EVM data.
- A reliable source of authoritative EVM data and access for The Office of the Secretary of Defense (OSD), the Services, and the DoD Components.
- Houses Contract Performance Reports (CPRs), Contract Funds Status Report (CFSR), and the Integrated Master Schedules (IMS) submitted by contractors

(and reviewed and approved by Program Management Offices) for ACAT 1C & 1D (MDAP) and ACAT 1A (MAIS) programs.

- Approximately 80 ACAT 1A, 1C, and 1D programs and 210 contracts and tasks reporting data (Defense Cost and Resource Center (DCARC), 2014).

Figure 14 provides a graphic representation of the EVM-CR (Defense Cost and Resource Center (DCARC), 2014). As discussed in the previous chapter, the primary EVM data of interest for schedule assessment are: Budget at Complete (BAC), program start date, the estimated completion date (ECD) for the program, Budgeted Cost of Work Performed (BCWP), Budgeted Cost of Work Scheduled (BCWS), and the Integrated Master Schedule (IMS).

The programs of interest were selected based on commodity and contract type: DoD space programs and development contracts. The commodity filter narrowed the results to thirteen initial programs listed in Table 11. The following three programs were removed because the EVM-CR did not contain development contracts for them: the Enhanced Polar System (EPS), Evolved Expendable Launch Vehicle (EELV), and National Polar-Orbiting Operational Environmental Satellite System (NPOESS). The next data criteria are completed contracts or contracts that were reported as 90% complete or greater. The 90% number was used as a benchmark for near complete because the Selected Acquisition Report (SAR) does not require contracts past 90% complete to report progress. As a result of these criteria, the following programs were eliminated: Family of Advanced Beyond Line-of-Sight Terminals (FAB-T), Global Positioning System III (GPS III), Global Positioning System Next Generation

Operational Control System (GPS OCX) Phase B, and Military GPS User Equipment (MGUE). The Advanced Extremely High Frequency Satellite (AEHF) and Space-Based Infrared System High Component (SBIRS HIGH) were included in this analysis because they were considered near complete at 99 and 96 percent complete. Table 12 shows the six programs and ten contracts that were analyzed.

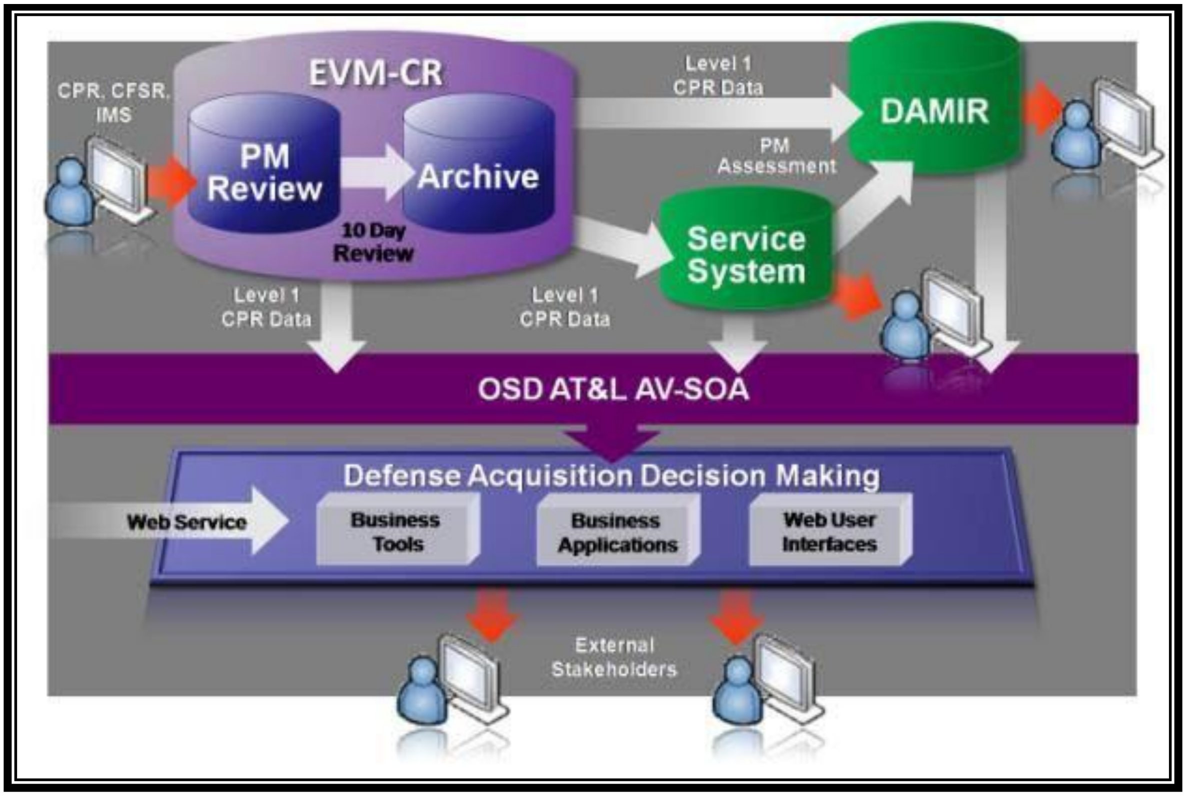


Figure 14: EVM Central Repository Overview

The contracts were classified as stable or unstable in an attempt to answer the research question, "are the forecasts accurate for contracts with Over-Target-Baselines (OTBs)?" Table 13 shows programs without an OTB while Table 14 lists programs with OTBs. Further analysis by system type (surveillance, communication, or navigation) was considered, but ultimately was not conducted because the dataset was already limited in size.

Table 11: Initial Space System Programs

Program Name	Number of Contracts	Development Contracts
Advanced Extremely High Frequency Satellite (AEHF)	2	1
Enhanced Polar System (EPS)	2	0
Evolved Expendable Launch Vehicle (EELV)	2	0
Family of Advanced Beyond Line-of-Sight Terminals (FAB-T)	6	1
Global Positioning System III (GPS III)	2	1
Joint Tactical Networks (JTN) - Army	5	1
Military GPS User Equipment (MGUE)	3	3
Mobile User Objective System (MUOS) - Navy	1	1
National Polar-Orbiting Operational Environmental Satellite System (NPOESS)	1	0
Navstar Global Positioning System (Navstar GPS)	4	3
Next Generation Operational Control System (GPS OCX)	3	3
Space-Based Infrared System High Component (SBIRS High)	5	1
Wideband Global SATCOM (WGS)	2	2
Total	38	17

Table 12: Contracts Analyzed

Program	Contract	Task	Data Points
Advanced Extremely High Frequency Satellite (AEHF)	F04701-02-C-0002	SDD	144
Mobile User Objective System (MUOS) - Navy	N00039-04-C-2009	CLIN 0400	55
Next Generation Operational Control System (GPS OCX)	FA8807-08-C-0001	System Design	21
Next Generation Operational Control System (GPS OCX)	FA8807-08-C-0003	System Design	24
Navstar Global Positioning System (Navstar GPS)	FA8807-06-C-0001	MUE	71
Navstar Global Positioning System (Navstar GPS)	FA8807-06-C-0003	MUE	68
Navstar Global Positioning System (Navstar GPS)	FA8807-06-C-0004	MUE	70
Space-Based Infrared System High Component (SBIRS High)	F04701-95-C-0017	RDT&E	212
Wideband Global SATCOM (WGS)	FA8808-06-C-0001	Blk 2	87
Wideband Global SATCOM (WGS)	FA8808-10-C-0001	B2FO	43

Table 13: Contracts without an OTB

Program	Contract
GPS OCX	FA8807-08-C-0001
GPS OCX	FA8807-08-C-0003
WGS	FA8808-06-C-0001
WGS	FA8808-10-C-0001

Table 14: Contracts with One or More OTB

Program	Contract	OTBs
AEHF	F04701-02-C-0002	3
MUOS	N00039-04-C-2009	3
NAVSTAR GPS	FA8807-06-C-0001	1
NAVSTAR GPS	FA8807-06-C-0003	4
NAVSTAR GPS	FA8807-06-C-0004	1
SBIRS HIGH	F04701-95-C-0017	4

Data Limitations

Although monthly CPRs are reviewed by the program management office prior to being entered into the EVM-CR, the data may contain inaccuracies. The data used in this analysis were reviewed for logic and accuracy. The key finding was missing data. For missing values, linear interpolation was used (prior reported value, next reported value, and the time elapsed between the two periods). The lists of missing data are located in Appendix A (Table 44 to Table 61).

Forecasting Method: EVM Index Based

The duration estimate is called the Time Estimate at Completion (TEAC). The index based TEACs have the following form:

Equation 37: Time Estimate at Completion (TEAC)

$$TEAC = IMS\ PD/PF$$

Where the IMS PD is the planned duration as reported in that month's IMS and PF is one of the earned value index performance factors. The IMS planned duration is calculated as follows: the days between the reported contract start date and the IMS completion date. The days are then converted to months. Table 15 lists the performance factors (PFs) that are used in this analysis. Time series performance factors are denoted by T.S. The SPI(t) metric was calculated with Lipke's earned schedule calculator from the Earned Schedule website (http://www.earnedschedule.com/Calculator.shtml).

Table 15: List of Performance Factors

Name	Static	Time Series
Baseline Execution Index	BEI	BEI (T.S.)
Schedule Performance Index	SPI	SPI (T.S.)
Cost Performance Index	CPI	CPI (T.S.)
Earned Schedule SPI	SPI(t)	SPI(t) (T.S.)
Schedule Cost Index	SPI*CPI	SPI (T.S.)*CPI (T.S.)
Schedule Cost Index (ES)	SPI(t)*CPI	SPI(t) (T.S.) *CPI (T.S.)
Enhanced Schedule Cost Index	BEI*CPI*SPI	BEI*CPI (T.S.)*SPI (T.S.)
Enhanced Schedule Cost Index (ES)	BEI*CPI*SPI(t)	BEI (T.S.)*CPI (T.S.)*SPI(t) (T.S.)
Enhanced CPI	BEI*CPI	BEI (T.S.)*CPI (T.S.)
Enhanced SPI	BEI*SPI	BEI (T.S.)*SPI (T.S.)
Enhanced SPI(t)	BEI*SPI(t)	BEI (T.S.)*SPI(t) (T.S.)

Forecasting Method: EVM Index Based plus Time Series Analysis

Time series analysis was conducted with JMP® 11.0 to estimate the CPI, SPI, SPI(t), and BEI parameters. The Box-Jenkins methodology for ARIMA models was used for this time series analysis. The Box-Jenkins methodology consists of three phases:

Identification, Estimation and Testing, and Application (Makridakis, Wheelwright, & Hyndman, 1998).

Initiating the Analysis

Prior to conducting the analysis, the number of autocorrelation lags and forecast periods must be determined. The number of autocorrelation lags will be n-1 until a maximum of 25 is reached. For example, the SPI(t) at month 20 will have 19 autocorrelation lags to calculate a forecasted SPI(t). Month 30 will use a maximum of 25 lags in the analysis. The number of forecast periods is one (the next period). With the autocorrelation lags and forecast periods determined we begin the analysis using the Time Series command in JMP® 11. The initial output of the analysis is a plot of the data as depicted in Figure 15.

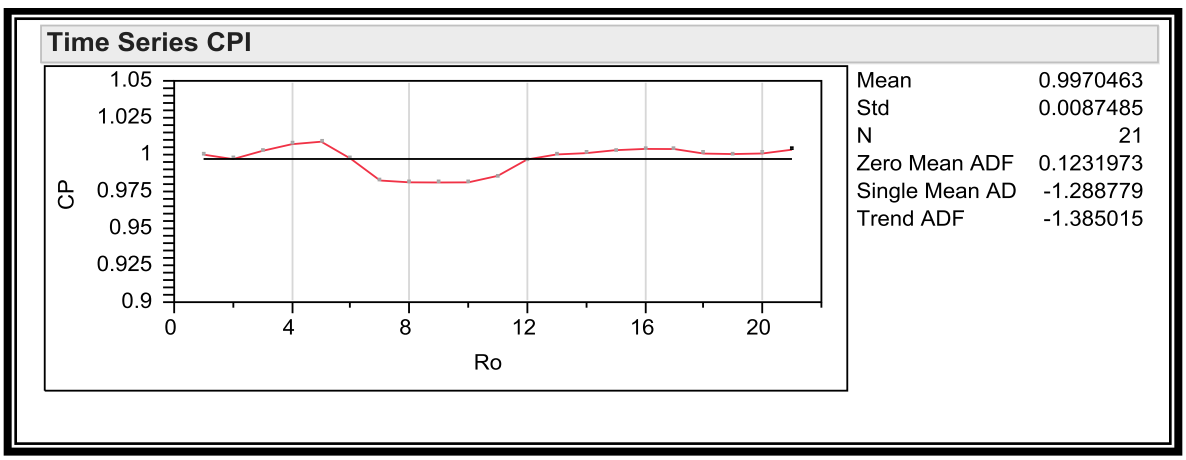
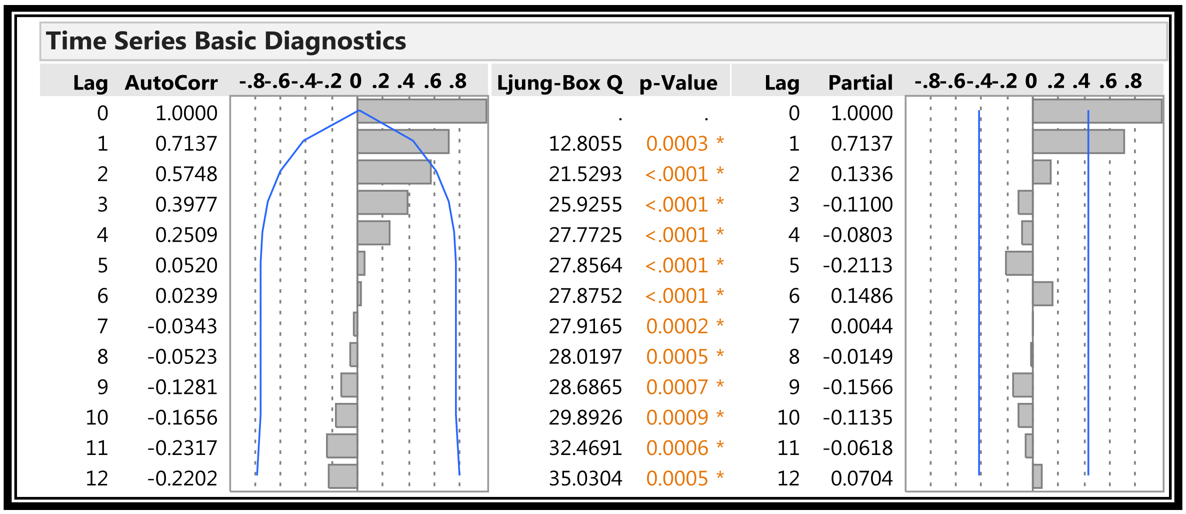


Figure 15: CPI Time Series Graph

Phase I - Identification

Data Preparation

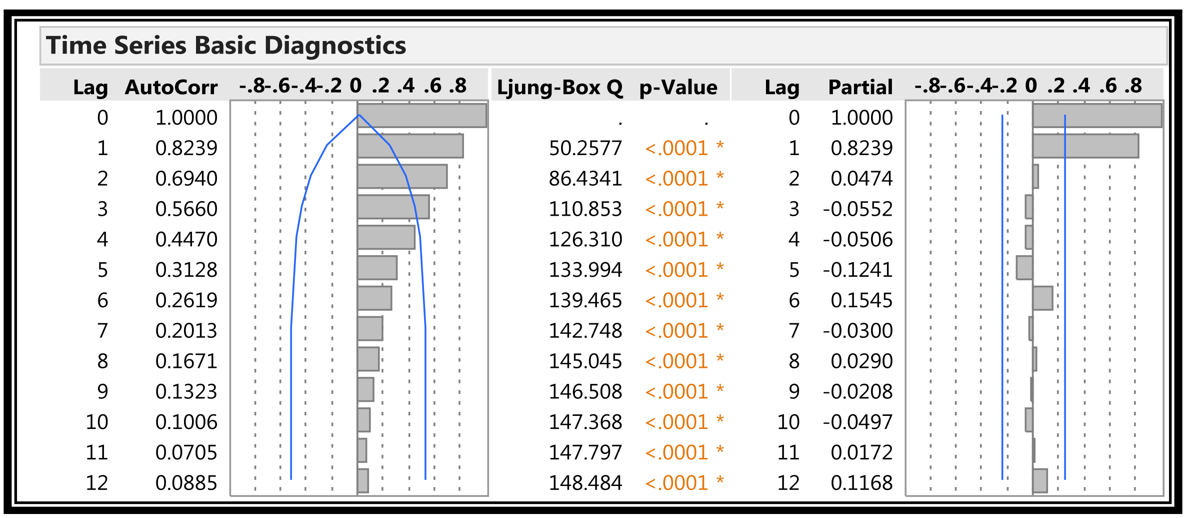
The analysis begins with an examination of the ACFs and PACF for stationarity. Figure 16 shows a stationary time series while Figure 17 shows a potential non-stationary time series. When a visual examination of the ACF graph does not provide conclusive results, the Augmented Dickey-Fuller test (ADF) can be used. The ADF test determines stationarity with a mathematical test. A negative value denotes a stationary time series. We can refer back to Figure 15 and conclude that this time series is stationary because single mean and trend ADFs are negative. If necessary, differencing can be used to remove non-stationarity in Phase II.



Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.7137	12.8055	0.0003 *	1	0.7137
2	0.5748	21.5293	<.0001 *	2	0.1336
3	0.3977	25.9255	<.0001 *	3	-0.1100
4	0.2509	27.7725	<.0001 *	4	-0.0803
5	0.0520	27.8564	<.0001 *	5	-0.2113
6	0.0239	27.8752	<.0001 *	6	0.1486
7	-0.0343	27.9165	0.0002 *	7	0.0044
8	-0.0523	28.0197	0.0005 *	8	-0.0149
9	-0.1281	28.6865	0.0007 *	9	-0.1566
10	-0.1656	29.8926	0.0009 *	10	-0.1135
11	-0.2317	32.4691	0.0006 *	11	-0.0618
12	-0.2202	35.0304	0.0005 *	12	0.0704

Figure 16: Plots of ACF and PACF (Stationary)



Time Series Basic Diagnostics

Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.8239	50.2577	<.0001 *	1	0.8239
2	0.6940	86.4341	<.0001 *	2	0.0474
3	0.5660	110.853	<.0001 *	3	-0.0552
4	0.4470	126.310	<.0001 *	4	-0.0506
5	0.3128	133.994	<.0001 *	5	-0.1241
6	0.2619	139.465	<.0001 *	6	0.1545
7	0.2013	142.748	<.0001 *	7	-0.0300
8	0.1671	145.045	<.0001 *	8	0.0290
9	0.1323	146.508	<.0001 *	9	-0.0208
10	0.1006	147.368	<.0001 *	10	-0.0497
11	0.0705	147.797	<.0001 *	11	0.0172
12	0.0885	148.484	<.0001 *	12	0.1168

Figure 17: Plots of ACF and PACF (potential non-stationary)

Model Selection

The model selection stage requires an examination of the time series graph, ACF, and PACF plots to identify potential models. Figure 16 shows a strong candidate for an autoregression (AR) model. The JMP® ARIMA Model Group function is an aid to the model selection process because it can be used to compare multiple models at once. As discussed in Chapter two, with the parameters from Table 6 we can produce twenty seven potential models (listed in Table 7). After each month of data is analyzed, the diagnostics are produced. Each of the twenty seven models from Table 7 will be considered. These models will be entered into JMP® ARIMA model group command which will generate an output similar to Table 16.

Phase II – Estimation and Testing

Estimation

Each model's usefulness is evaluated by the Akaike Information Criterion (AIC). Lower AIC values are associated with a better model (Makridakis, Wheelwright, & Hyndman, 1998). In this analysis, the model with the lowest average AIC is deemed the best model and a candidate to forecast the performance factor. However, a diagnostics check of the residuals must be conducted prior to using the model for forecasting.

Diagnostics

As previously discussed, in order for a forecasting model to be considered adequate, the residuals should be white noise. Figure 18 shows this model's residuals are from white noise because they are all within the range denoted by the blue line (alpha = 0.05). A more robust test is the Ljung-Box Q portmanteau test of residuals. At an alpha of 0.05, all of the values are not significant; therefore the residuals can be considered a white noise series. If the model residuals are not considered white noise, then we will return to model selection stage and start the process again. If the model residuals are from a white noise series we can proceed to forecasting with the model.

Table 16: Time Series Model Comparison

Model	DF	Variance	AIC	SBC	R Square	-2LogLH	Weights	MAPE	MAE
AR(1)	69	0.00151	-256.39	-251.86	0.681	-260.39	0.6125	2.145	0.019
ARMA(1, 1)	68	0.00153	-254.69	-247.90	0.682	-260.69	0.2616	2.178	0.019
ARIMA(1, 1, 1)	67	0.00147	-252.48	-245.73	0.676	-258.48	0.0866	2.312	0.021
IMA(1, 1)	68	0.00162	-248.83	-244.33	0.658	-252.83	0.0139	2.267	0.021
ARI(1, 1)	68	0.0016	-248.68	-244.18	0.657	-252.68	0.0129	2.236	0.020
I(1)	69	0.00165	-248.57	-246.32	0.651	-250.57	0.0122	2.235	0.020
MA(1)	69	0.00254	-220.27	-215.74	0.452	-224.25	0	3.480	0.031
ARIMA(0, 0, 0)	70	0.00468	-178.35	-176.10	0	-180.35	0	4.986	0.045

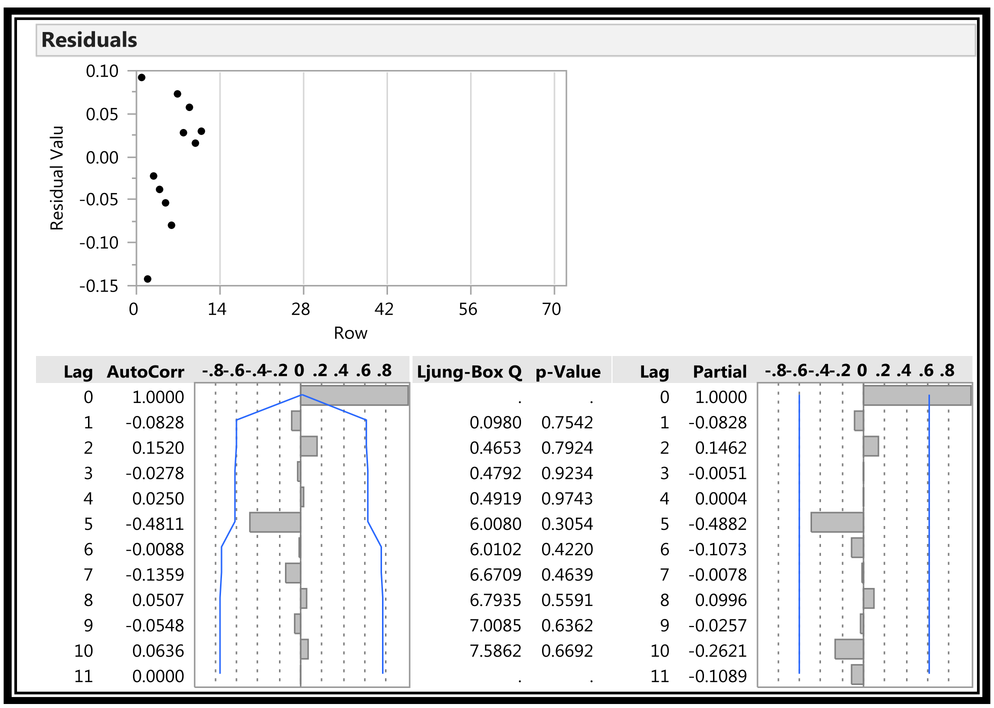


Figure 18: Plots of ACF and PACF for Residuals

Phase III – Application

Forecasting

Because of limited data in the early periods, time series forecasts will not be used until month four. For the first month, the reported value of the performance factor will be used. For the second month, the average of months one and two will be used. For the third month, the average of months one, two, and three will be used as the forecasted performance factor. From month four going forward, we used the forecasting model selected in Phase II with a maximum of twenty-five lags.

A fifty month contract should have forty-seven time series forecast values each for the index values (excluding months 1-3). These forecasted index values will be used as performance factors (PF) in the time estimate at completion (TEAC = IMS PD/PF) for that period.

Forecasting Method: Linear Regression

As discussed in Chapter Two's linear regression section, this method regresses the BCWP against time (months). The BAC is also regressed against time (months). The regressions are calculated from month three until the last reported month for each contract. For each monthly forecast, the next step is setting BCWP and BAC regression equations equal to each other to solve for the unknown month as displayed in Equation 38. After the intermediate calculation, the duration formula is simplified to Equation 39. If the BAC changed by more than 10% from one period to the next the analysis is reset. This means the analysis starts anew, the previous data points are not included in the regression calculations going forward. This approach helps smooth the forecast when large changes in BAC occur from one period to the next.

Equation 38: Regression Forecast (Intermediate Calculation)

$$\text{BCWP intercept} + \text{BCWP coefficient} * \text{Months} = \text{BAC intercept} + \text{BAC coefficient} * \text{Months}$$

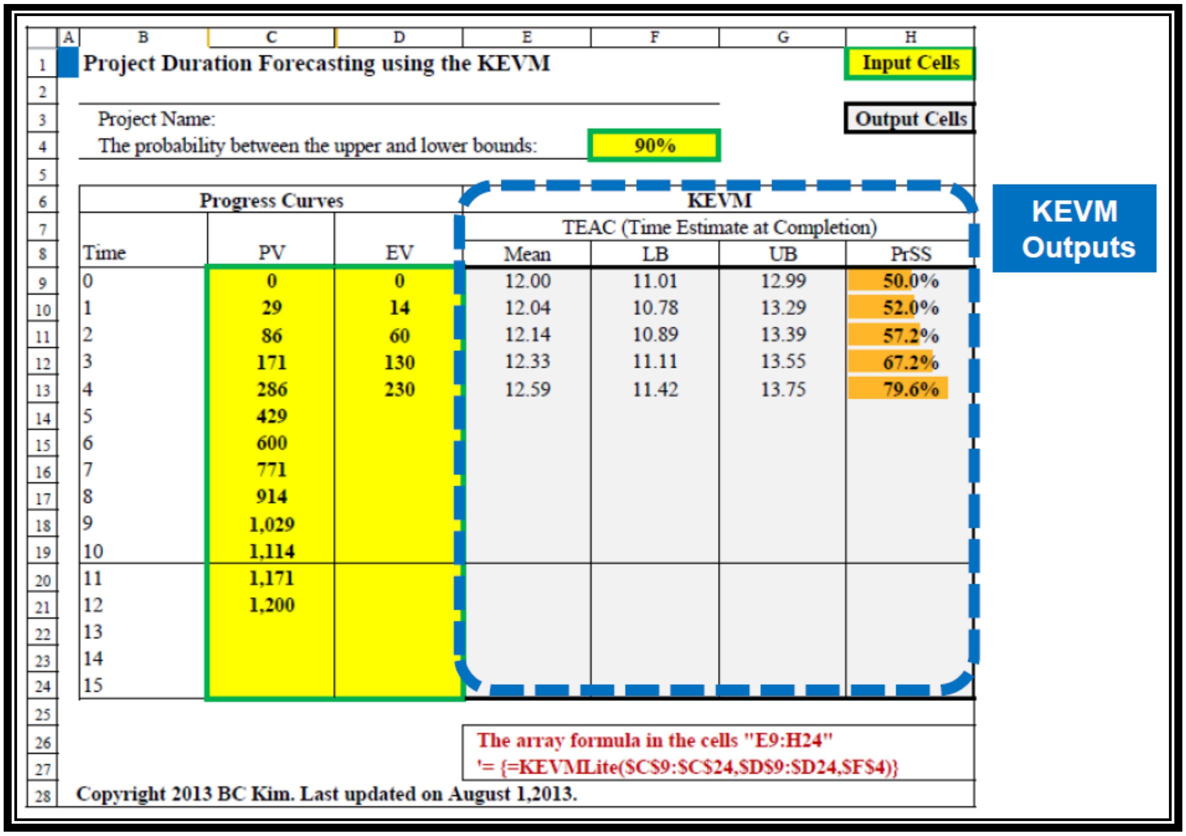
Equation 39: Duration Forecast (Regression Based)

$$\text{Months} = \frac{\text{BAC intercept} - \text{BCWP intercept}}{(\text{BCWP coefficient} - \text{BAC coefficient})}$$

Forecasting Method: Kalman Filter Forecast Method

The Kalman Filter Forecast Method was applied with the Excel tool KEVM Lite© developed by Kim (2010). The planned duration, the time phased planned values (also called the performance measurement baseline (PMB)), and the confidence level are the inputs required for this method. The confidence level is a decision variable; 95% was used in this analysis. The planned duration is based on the reported Estimated Completion Date (ECD). Portions of the PMB must be estimated if the monthly PMB is not known. The time phasing of the planned values is developed with linear interpolation of the reported BAC and planned duration.

After making the appropriate adjustments, the KEVM Lite © updates each month's forecast. This forecast contains a mean, upper bound (UB), and a lower bound (LB) for the time estimate at completion (TEAC). In addition to the three TEAC estimates, the probability of schedule slip (PrSS) was calculated. Examples of the TEAC estimates and PrSS are displayed in Figure 19; the mean value was used in this analysis.



Project Duration Forecasting using the KEVM

Input Cells

Output Cells

Project Name:

The probability between the upper and lower bounds: 90%

	Progress Curves		KEVM			
			TEAC (Time Estimate at Completion)			
Time	PV	EV	Mean	LB	UB	PrSS
0	0	0	12.00	11.01	12.99	50.0%
1	29	14	12.04	10.78	13.29	52.0%
2	86	60	12.14	10.89	13.39	57.2%
3	171	130	12.33	11.11	13.55	67.2%
4	286	230	12.59	11.42	13.75	79.6%
5	429					
6	600					
7	771					
8	914					
9	1,029					
10	1,114					
11	1,171					
12	1,200					
13						
14						
15						

The array formula in the cells "E9:H24"

'= {=KEVMLite($C$9:$C$24,$D$9:$D24,$F$4)}

Copyright 2013 BC Kim. Last updated on August 1,2013.

Figure 19: KEVM Lite ©

Forecasting Method: Independent Duration Estimate (IDE)

The final technique used in this analysis was derived from Lofgren's research (2014). The IMS planned duration will be modified and used with the performance factors to calculate an Independent Duration Estimate (IDE). The schedule slip metric will be calculated with the formula in Equation 40 (Lofgren, 2014). Each unfinished task is considered for the schedule slip. As tasks are completed they are removed from consideration. The results for one example contract are displayed in Table 17. In this example, 4.2 months are added to the Integrated Master Schedule (IMS) planned duration of 49.1 months, for a total of 53.3 months. This schedule slip is added to the current planned duration to obtain an independent duration estimated (IDE) as shown in Equation

43. The IDE will be used with the performance factors to calculate a TEAC. The following equations are used to calculate the parameters in Table 17. These equations were previously listed in chapter two, they are listed again for clarity and convenience.

Equation 40: Schedule Slip

Slip = Max (Current Finish Date – Baseline Finish Date – Total Slack)

Equation 41: IMS Planned Duration

IMS planned duration = start date from CPR to IMS reported end date

Equation 42: CPR Planned Duration (status quo)

CPR PD = start date from CPR to Estimated Completion Date from CPR

Equation 43: Independent Duration Estimate

Independent Duration Estimate = IMS planned + schedule slip estimate

Equation 44: Enhanced IDE

Enhanced IDE = IDE/PF

Table 17: IMS Analysis (Current Month Compared to Baseline)

Task Name	Baseline Finish (IMS#1) 4/15/08	Baseline Total Slack	Current Finish (IMS#2) 5/20/08	Finish Variance (days)	Slip (days)	Slip (months)
ASIC Build 1-2-3-4 Integrat.	01/30/08	9	05/02/08	92	83	2.77
PSP Develop Test Cases 1	06/02/08	-47	06/02/08	0	47	1.57
IO : Det Design (S2) Ph 1	05/16/08	-80	07/02/08	46	126	4.20
				MAX	126	**4.20**

Finally, if lapses in data occur the IMS PD will be used for the IDE (see Appendix A). Lapses occurred most frequently in the beginning of the contract.

Evaluating the Forecasting Models (Accuracy, Timeliness, and Reliability)

In order to determine the usefulness of the forecasting models an evaluation measure must be selected. The evaluation measure used in this research is the Mean Absolute Percent Error (MAPE). There are many forecasting evaluation measures, but the MAPE is arguably the easiest to explain and understand. The MAPE formula is exhibited in Equation 45 (Makridakis, Wheelwright, & Hyndman, 1998). In this equation, *n* equals the total number of observations (months) and *t* equals the time of the forecast.

Equation 45: Mean Absolute Percentage Error (MAPE)

$$\text{Mean Absolute Percentage Error (MAPE} = \frac{1}{\text{n}}\sum \text{Abs}[(\text{Actual}_\text{t} - \text{Forecast}_\text{t})/\text{Actual}_\text{t}]$$

Models with lower MAPE values (closer to zero) are more accurate. For example, a MAPE of 0% represents a perfect forecast. A MAPE of 15% means that the forecast is underestimating or overestimating the true value by 15% on average. Figure 20 displays one model's [IDE / (SPI(t) (T.S.) * BEI] forecast compared to the status quo forecast (CPR PD); the IDE based forecast is more accurate than the status quo until the late stage of the program (80% to 100%). Additionally, in order to assess the timeliness, the MAPE will be calculated in 10% intervals from 0% to 100%.

Table 18 compares six models to the planned duration using the previously discussed metrics; partial results are displayed because of space constraints (43 models). For individual contracts the following forecast models are reported: the CPR PD, IMS PD, IDE, most accurate IMS PD/PF, most accurate IDE/PF, Regression, and Kalman filter method.

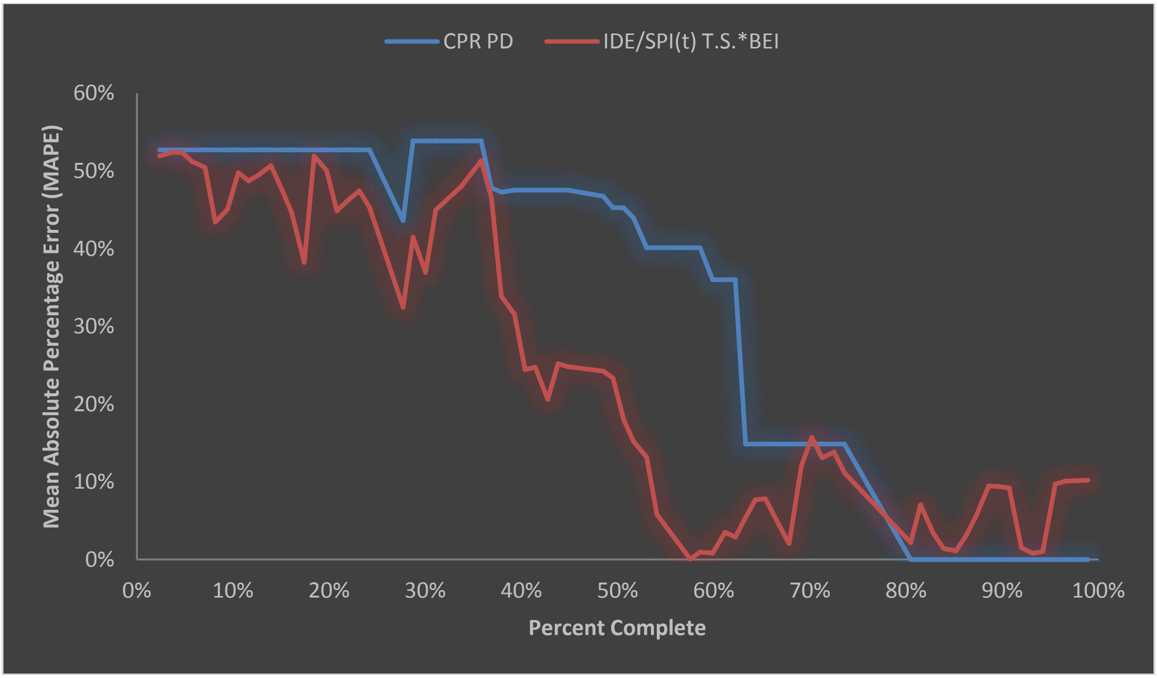


Figure 20: Duration MAPE over Time

Table 18: Forecast Model Intervals and Overall MAPE

Percent Complete Interval	**Forecasting Model**						
	CPR PD (status quo)	IMS PD	IDE	IMS PD / [SPI(t) * CPI * BEI]	IDE / [SPI(t) (T.S.) * BEI]	Regress-ion	Kalman Filter
0 to 10	52.72%	52.72%	52.72%	37.76%	49.55%	74.58%	52.72%
11 to 20	52.72%	52.72%	52.72%	42.05%	47.91%	80.11%	52.72%
21 to 30	51.75%	51.75%	51.75%	43.07%	42.10%	63.11%	48.86%
31 to 40	50.26%	50.45%	43.34%	42.26%	40.10%	52.74%	52.42%
41 to 50	47.04%	46.95%	29.00%	36.40%	23.83%	52.29%	46.07%
51 to 60	40.82%	41.84%	17.38%	21.41%	7.72%	53.17%	44.53%
61 to 70	19.57%	19.57%	14.61%	7.03%	6.86%	50.60%	35.93%
71 to 80	11.16%	11.16%	11.16%	5.03%	10.06%	40.89%	27.36%
81 to 90	0.00%	0.00%	8.32%	6.78%	5.07%	15.14%	0.71%
91 to 100	0.00%	0.00%	4.33%	5.56%	6.08%	15.79%	1.20%
MAPE	33.05%	33.16%	29.26%	25.14%	24.45%	50.57%	36.44%

In addition to reporting the results of individual contracts, results are grouped into OTB versus non-OTB contracts (631 and 175 observations). The analysis is further grouped by long duration (SBIRS and AEHF), medium duration (MUOS, NAVSTAR GPS, and WGS), and short duration contracts (GPS OCX). The long duration group has 356 observations, the medium duration group has 405 observations, and the short duration group has 45 observations. The analysis is further categorized to contracts with the data necessary to create an IDE (7 of 10 contracts with 617 observations). The IDE models will be compared to the other model types within the same data set (seven contracts). The last grouping is an aggregate of forecasts across all contracts (this does not include IDE models because three contracts did not have available data); in this analysis there are 806 total forecasts for each model. Finally, due to the potential for similar accuracy results the models were analyzed with the Tukey-Kramer HSD multiple comparison of means function via JMP®. The purpose of this test is to determine if the means of the absolute percent errors (APEs) are significantly different from each other and different from the status quo. The Tukey-Kramer HSD uses pooled variances; therefore, before proceeding we must determine if the variances are equal (JMP, 2013).

Test for Unequal Variances: Levene Test

We tested for unequal variances using the Levene's test with an alpha of 0.05 and the following hypothesis:

- $\mathbf{H_0}$: the variances are the same: $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_k^2$
- $\mathbf{H_a}$: at least one variance is different

If the p-value is greater than alpha we fail to reject the null and conclude the variances are equal. If the p-value is less than alpha we reject the null and conclude at least one variance is different (JMP®, 2013).

Multiple Comparisons of Means: Tukey-Kramer HSD

We can use the Tukey-Kramer HSD method to compare means if the APEs are normally distributed (or the number of observations are greater than 30) and the variances are equal. An alpha of 0.05 is used unless otherwise noted. If the APEs are not normally distributed (or the number of observations are less than 30) or the variances are not equal it is recommend to use an alternative method.

Summary

This chapter described how the forecasting models were developed. A description of the data source, data selected and its limitations was provided. Next, we discussed the systematic approach to compute the status quo (CPR PD), EVM Index Performance Factors, EVM Index Performance Factors (Time Series based), linear regression, the Kalman Filter Forecast Method, and the Independent Duration Estimate (IDE). In summary, this research utilizes five types of forecasting techniques:

1. CPR PD (status quo)
2. IMS PD and Enhanced IMS PD = IMS PD/PF (non-time series and time series)
3. Linear Regression (Smoker, 2011)
4. Kalman Filter Forecasting Method (Kim, 2007 & 2010)
5. IDE (IDE = IMS PD + Schedule Slip) and Enhanced IDE = IDE/PF (non-time series and time series) (Lofgren, 2014)

The status quo is the base case and serves as a comparison for the relative accuracy of the other techniques. The Kalman Filter and Regression methods are standalone techniques in this research and the results are easy to distinguish. The IMS PD and IDE are similar because they both use the planned duration from the IMS plus the performance factors. The distinguishing factor is the schedule slip metric in the IDE. Time series analysis was not a standalone model, but an addition to both the IMS PD and the IDE performance factors (PF). Models with time series performance factors are denoted by T.S. For example the model IMS PD/ [SPI(t)*BEI (T.S.)] has a BEI time series performance factor. Finally, the model evaluation criterion was listed (MAPE) and the Tukey-Kramer HSD method was explained. In the next chapter, the results of this analysis are reported.

IV. Results and Discussion

Chapter Overview

In this chapter we review the research objective and investigative questions before reporting the accuracy of the schedule forecasting methods. The objective is to evaluate forecasting methods for space program duration based on the following criteria: accuracy, reliability, and timeliness. In support of the overarching research objective, the following questions were investigated:

1. What are the appropriate methods to estimate a program's duration?
2. How should accuracy be measured and how accurate are the various schedule estimating methods (individual contract, overall, and by various groupings)?
3. At what point in time (if at all) are the techniques more accurate than the status quo?
4. Are the forecasts accurate for programs with one or more over target baseline (OTB)?

The first question was exploratory in nature. Several forecasting methods were studied, the strengths and weakness of the various models were discussed in chapters two and three. The remaining questions comprise the bulk of the analysis; this chapter is dedicated to answering these questions.

Forecast Model Accuracy Results

All Contracts (No IDE Models)

Table 19 lists the MAPE for each model for the entire data set (806 observations). This does not include Independent Duration Estimate (IDE) models. The most accurate model across the entire data set is an improvement of 2.93% over the status quo (26.14% vs. 23.22%). With the exception of the regression approach (36.43%), each of the models lie within a narrow range (23.22% to 26.14%).

Table 19: MAPE - All Contracts (No IDE Models)

Forecasting Model	MAPE
IMS PD/ [SPI(t) (T.S.)*BEI]	23.22%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	23.25%
IMS PD/ [SPI(t) (T.S.)]	24.30%
IMS PD/ [SPI*CPI*BEI(T.S.)]	24.50%
IMS PD/ [SPI*CPI*BEI]	24.52%
IMS PD/ [SPI(t)]	24.59%
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	24.66%
IMS PD/ [SPI(t)*CPI*BEI]	24.75%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	24.84%
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)]	24.87%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	24.89%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	25.06%
IMS PD/ [SPI(t)*CPI]	25.07%
IMS PD/ [SPI(T.S.)]	25.11%
IMS PD/ [SPI(t)*CPI(T.S.)]	25.14%
IMS PD/ [SPI(t)(T.S.)*CPI]	25.20%
IMS PD/ [SPI*CPI]	25.25%
IMS PD/ [SPI(T.S.)*CPI]	25.26%
IMS PD/ [SPI]	25.34%
IMS PD	25.77%
Kalman Filter	25.94%
CPR PD (status quo)	26.14%
Regression	36.43%

Every model except regression was more accurate than the status quo. However, because many of the values were clustered together we conducted a Tukey-Kramer HSD analysis of means. Analyzing all of the models at once resulted in unequal variances. In chapter three we discussed the necessity of equal variances before we could use the Tukey-Kramer HSD method. We truncated the analysis to include the CPR PD and the most accurate models. The Levene test p-value was 0.9624, denoting equal variance (see Appendix B, Figure 36). The results of the Tukey-Kramer analysis are displayed in Figure 21; examining the connecting letters report from top to bottom, the models that do not have a letter in common are significantly different. Two models are significantly different from the status quo: [IMS PD/ SPI(t) (T.S.)*BEI(T.S.)] and [IMS PD/ SPI(t) (T.S.)* BEI]. These models are outlined with a blue box at the bottom of Figure 21.

Comparisons for all pairs using Tukey-Kramer HSD

Connecting Letters Report

Level			**Mean**
CPR PD (Status Quo)	A		0.26143040
IMS PD	A	B	0.25765533
IMS PD/SPI(t) T.S.*BEI(T.S.)*CPI(T.S	A	B	0.24846278
IMS PD/SPI(t)*CPI*BEI(T.S.)	A	B	0.24637221
IMS PD/SPI(t)	A	B	0.24586576
IMS PD/SPI*CPI*BEI(T.S.)	A	B	0.24476141
IMS PD/SPI(t) (T.S.)	A	B	0.24297630
IMS PD/SPI(t) (T.S.)*BEI (T.S.)		B	0.23233908
IMS PD/SPI(t) (T.S.)*BEI		B	0.23216898

Levels not connected by same letter are significantly different

Figure 21: Tukey-Kramer HSD - All Contracts

When evaluating all contracts we can say the two models are more accurate than the status quo and the difference is not likely to be random. The SPI(t) metric appears in both models reaffirming the research by Henderson (2004), Lipke (2004 & 2009), Vandevoorde and Vanhoucke, (2006), and Crumrine (2013). Additionally, each of the models had at least one time series based performance factor. Finally, the BEI appears in both of the models. The BEI did not appear in the forecasting literature, nevertheless these results suggest it is a valuable duration forecasting parameter.

IDE Data Set (includes 7 of 10 contracts)

Table 20 shows the results of the analysis of the seven contracts with IDE data. The two GPS OCX contracts and the AEHF contract did not have the IMS data suitable for developing IDEs. For this analysis, the most accurate model exhibits an improvement of 5.2% over the status quo (26.47% vs. 21.27%). Thirty-seven of the forty-three models are more accurate than the status quo. The seven most accurate models are IDE based. These results suggest Lofgren's approach (IDE) is the most accurate technique in this research. With the exception of regression (38.36%), the results fall within a range from 21.27% to 27.21%. Once again, many of the models were clustered. Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate models. The Levene test p-value was 0.3554, denoting equal variance (see Appendix B, Figure 37). We conducted a Tukey-Kramer HSD comparison of means to determine if the means were significantly different from each other and the status quo. The results of this analysis are displayed in Figure 22; examining the connecting letters report from top to bottom, eight models were significantly different from the CPR PD (status quo). These models are outlined with a blue box at the bottom of Figure 22. When evaluating the contracts with IDE data we can conclude that these models are more accurate than the status quo and the difference is not likely to be random. One model [IMS PD/ SPI(t) (T.S.)* BEI] identified as significantly from the all contracts data set also appears here.

Table 20: MAPE – IDE Data Set (includes 7 of 10 contracts)

Forecasting Model	MAPE
IDE/ [SPI (T.S.)]	21.27%
IDE/ [SPI(t)]	21.35%
IDE/ [SPI(t) (T.S.)]	21.40%
IDE/ [SPI]	21.50%
IDE/ [SPI(t) (T.S.)*BEI]	21.87%
IDE/ [SPI(t) (T.S.)*BEI (T.S.)]	21.89%
IDE	22.21%
IMS PD/ [SPI(t) (T.S.)*BEI]	22.95%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	22.98%
IMS PD/ [SPI(t) (T.S.)]	24.23%
IDE/ [SPI(T.S.)*CPI]	24.50%
IDE/ [SPI*CPI]	24.51%
IDE/ [SPI(T.S.)*CPI(T.S.)]	24.53%
IMS PD/ [SPI(t)]	24.60%
IMS PD/ [SPI*CPI*BEI(T.S.)]	25.01%
IMS PD/ [SPI*CPI*BEI]	25.06%
IMS PD/ [SPI(T.S.)]	25.21%
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	25.30%
IDE/ [SPI*CPI*BEI(T.S.)]	25.34%
IDE/ [SPI*CPI*BEI]	25.36%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	25.41%
IMS PD/ [SPI(t)*CPI*BEI]	25.43%
IMS PD/ [SPI]	25.52%
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)]	25.57%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	25.62%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	25.62%
IMS PD/ [SPI(t)*CPI]	25.72%
IDE/ [SPI(t)*CPI]	25.78%
IMS PD/ [SPI(t)*CPI(T.S.)]	25.79%
IMS PD/ [SPI*CPI]	25.89%
IMS PD/ [SPI(T.S.)*CPI]	25.89%
IMS PD/ [SPI(t)(T.S.)*CPI]	25.89%
IDE/ [SPI(t)(T.S.)*CPI]	25.92%
IMS PD	25.94%
Kalman Filter	25.95%
IDE/ [SPI(t)*CPI(T.S.)]	25.95%
IDE/ [SPI(t)(T.S.)*CPI(T.S.)]	26.16%
CPR PD (status quo)	26.47%
IDE/ [SPI(t)*CPI*BEI(T.S.)]	26.75%
IDE/ [SPI(t)*CPI*BEI]	26.78%
IDE/ [SPI(t) (T.S.)*BEI(T.S.)*CPI(T.S.)]	27.19%
IDE/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	27.21%
Regression	38.36%

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Connecting Letters Report

Level				Mean
CPR PD (Staus Quo)	A			0.26468088
Kalman Filter	A	B		0.25946629
IMS PD	A	B		0.25940130
IDE/SPI*CPI	A	B	C	0.24507131
IMS PD/SPI(t) (T.S.)	A	B	C	0.24227828
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	A	B	C	0.22977423
IMS PD/SPI(t) (T.S.)*BEI		B	C	0.22945818
Independent Duration Estimate (IDE			C	0.22208849
IDE/SPI(t) (T.S.)*BEI (T.S.)			C	0.21894668
IDE/SPI(t) (T.S.)*BEI			C	0.21872853
IDE/SPI			C	0.21503841
IDE/SPI(t) (T.S.)			C	0.21397974
IDE/SPI(t)			C	0.21350324
IDE/ SPI(T.S.)			C	0.21265981

Levels not connected by same letter are significantly different

Figure 22: Tukey-Kramer HSD - IDE Data Set (7 out of 10 contracts)

Non OTB Group (GPS OCX and WGS)

Table 21 lists the MAPE for each model for contracts without an OTB (GPS OCX and WGS). This does not include IDE models. OTB and non-OTB contracts were not compared for the IDE analysis because of the limited dataset (2 non-OTBs and 5 OTBs). The most accurate model is an improvement of 2.17% over the status quo (25.50% vs. 23.33%). The range is relatively narrow, from 23.33% to 27.79%. The two models from the all contracts analysis are also the most accurate here: IMS PD/ [SPI(t) (T.S.) *BEI (T.S.)] and IMS PD/ [SPI(t) (T.S.)* BEI]. Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate model. The Levene test p-value was 0.1302, denoting equal variance (see Appendix B, Figure 38). Next, we conducted a Tukey-Kramer HSD comparison of means. According to the connecting letters report (Figure 23) the model was not significantly different from the status quo. Therefore we cannot conclude that any of these models are better than the status quo when forecasting the duration of non-OTB contracts (alpha = 0.05). One model [IMS PD/ SPI(t) (T.S.)* BEI] becomes statistically different than the status quo if the alpha level is relaxed (alpha = 0.15) (Figure 24).

Table 21: MAPE - Non OTB Group (4 Contracts & 175 Observations)

Forecasting Model	MAPE
IMS PD/ [SPI(t) (T.S.)*BEI]	23.33%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	23.57%
IMS PD/ [SPI(t) (T.S.)]	23.78%
IMS PD	24.35%
IMS PD/ [SPI(T.S.)]	24.41%
IMS PD/ [SPI(t)(T.S.)*CPI]	24.55%
IMS PD/ [SPI(t)]	24.77%
IMS PD/ [SPI(t)*CPI]	24.79%
IMS PD/ [SPI(t)*CPI*BEI (T.S.)]	24.84%
IMS PD/ [SPI(t)*CPI*BEI]	24.87%
IMS PD/ [SPI(t)*CPI (T.S.)]	24.93%
IMS PD/ [SPI]	25.18%
IMS PD/ [SPI(t) (T.S.)*CPI (T.S.)]	25.31%
IMS PD/ [SPI(t) T.S.*BEI (T.S.)*CPI (T.S.)]	25.31%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI (T.S.)]	25.34%
Kalman Filter	25.38%
IMS PD/ [SPI*CPI*BEI (T.S.)]	25.43%
IMS PD/ [SPI*CPI*BEI]	25.46%
IMS PD/ [SPI (T.S.)*CPI (T.S.)]	25.46%
CPR PD (status quo)	25.50%
IMS PD/ [SPI*CPI]	25.95%
IMS PD/ [SPI(T.S.)*CPI]	26.19%
Regression	27.79%

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.96680	0.05

Connecting Letters Report

Level			Mean
CPR PD (status quo)	A		0.25493771
IMS PD/ [SPI(t) (T.S.)*BE	A		0.23333600

Levels not connected by same letter are significantly different

Figure 23: Tukey-Kramer HSD - No OTB

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.44272	0.15

Connecting Letters Report

Level			Mean
CPR PD (status quo)	A		0.25493771
IMS PD/ [SPI(t) (T.S.)*BE		B	0.23333600

Levels not connected by same letter are significantly different

Figure 24: Tukey-Kramer HSD - No OTB (alpha = 0.15)

The difference is not as pronounced as the previous analysis and is more susceptible to type I error (false positive). Why are the models less accurate for non-OTB contracts? Possible explanations include: the schedule performance is more stable for short and non-OTB contracts. The three shortest duration contracts were in this analysis (25.0, 28.3, and 47.4 months). Another possible explanation is lower cost and schedule growth for the non-OTB contracts. The four non-OTB contracts had an average schedule growth of 60.8% (median = 59.2%) compared to 135.8% (median = 94.8%) for the six OTB contracts. Additionally, the four non-OTB contracts had an average cost growth of 47.9% (median = 19.4%) compared to 170.5% (median = 147.8%) for the six OTB contracts. Low schedule and cost growth may indicate better initial schedule estimates and impact from management decisions. Therefore, there should be less room for accuracy improvement over the status quo estimate. Short duration contracts may be less uncertain than lengthier contracts because there is less time for changes and other unforeseen issues. Contract length, OTBs, schedule growth, and cost growth are further explored in the subsequent sections.

OTB Group (6 Contracts & 631 Observations)

Table 22 lists the MAPE for each model for contracts with at least one OTB. This does not include IDE models. In this grouping the most accurate model is an improvement of 3.16% over the status quo (26.32% vs. 23.16%). With the exception of regression (38.83%) the models lie within a narrow range (23.16% to 26.32%). Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate models. The Levene test p-value was 0.1305, denoting equal variance (see Appendix B, Figure 39). Next, we conducted a Tukey-Kramer HSD comparison of means. The connecting letters report (Figure 25) shows two models that are significantly different than the status quo: IMS PD/ [SPI(t) (T.S.)*BEI] and IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]. Both of the models are more accurate than the status quo and have been among the most accurate models for each type of analysis.

Table 22: MAPE – OTB Group (6 Contracts & 631 Observations)

Forecasting Model	MAPE
IMS PD/ [SPI(t) (T.S.)* BEI (T.S.)]	23.16%
IMS PD/ [SPI(t) (T.S.)* BEI]	23.18%
IMS PD/ [SPI* CPI* BEI(T.S.)]	24.24%
IMS PD/ [SPI* CPI* BEI]	24.26%
IMS PD/ [SPI(t) (T.S.)]	24.44%
IMS PD/ [SPI(t)]	24.53%
IMS PD/ [SPI(t)* CPI* BEI(T.S.)]	24.61%
IMS PD/ [SPI(t) (T.S.)* CPI(T.S.)]	24.70%
IMS PD/ [SPI(t)* CPI* BEI]	24.71%
IMS PD/ [SPI(t) (T.S.) *BEI (T.S.)* CPI(T.S.)]	24.74%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	24.77%
IMS PD/ [SPI (T.S.)*CPI (T.S.)]	24.94%
IMS PD/ [SPI (T.S.)* CPI]	25.00%
IMS PD/ [SPI* CPI]	25.06%
IMS PD/ [SPI(t)* CPI]	25.15%
IMS PD/ [SPI(t)* CPI(T.S.)]	25.19%
IMS PD/ [SPI (T.S.)]	25.30%
IMS PD/ [SPI(t) (T.S.)* CPI]	25.38%
IMS PD/ [SPI]	25.39%
Kalman Filter	26.10%
IMS PD	26.16%
CPR PD (status quo)	26.32%
Regression	38.83%

Why is the accuracy improvement significant for contracts with OTBs, but not non-OTB contracts? Contracts that undergo OTBs may have done so because the original estimates were overly optimistic. The hypothesis is contracts with OTBs have more potential for improved accuracy (over the status quo estimate). This relationship will be examined further in the subsequent sections.

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
2.72941	0.05

Connecting Letters Report

Level			Mean
CPR PD (status quo)	A		0.26323106
IMS PD/ [SPI*CPI*BEI]	A	B	0.24257623
IMS PD/ [SPI*CPI*BEI (T.S.)]	A	B	0.24210365
IMS PD/ [SPI(t) (T.S.)*BEI]		B	0.23184532
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.		B	0.23139699

Levels not connected by same letter are significantly different

Figure 25: Tukey-Kramer HSD (OTB)

Individual Contracts

We have examined the forecasting models with a variety of groupings. A handful of models have consistently appeared as the most accurate. Is there a single model that dominates on the individual contract level? Table 23 lists the most accurate model for each of the ten contracts along with the status quo model to illustrate the accuracy improvement. Detailed accuracy results for each contract are listed in Appendix C, beginning with Table 62 and Figure 43. Not surprisingly, no single model is the most accurate for each contract. In fact, the same model was not the most accurate for any two contracts. Of course similarities exist between the models and their parameters. Of note, models with SPI(t) are among the most accurate in 7 of 10 contracts. Models with BEI are among the most accurate in 5 of 10 contracts. Time series performance factors appear in 6 of the 10 most accurate models. IDE based models are the most accurate in 6 out of 7 contracts where data were available. These results reinforce the previous analysis. SPI(t) is a consistent performance factor for duration forecasting. BEI is not as strong, but has displayed validity in this research. Time series analysis can further enhance the index based models. Finally, the IDE approach has routinely been the most accurate for contracts with the available IMS data.

Table 23: Most Accurate Model by Contract

Program	Contract	Final Duration	Model	CPR PD (Status Quo)	Best Model	Delta
GPS OCX	FA8807-08-C-0001	25.0	[IMS PD/ SPI(t) (T.S).* BEI (T.S.)* CPI(T.S.)]	20.41%	18.37%	2.04%
GPS OCX	FA8807-08-C-0003	28.3	[IMS PD/ SPI(t)*CPI* BEI]	25.71%	21.98%	3.73%
WGS	FA8808-06-C-0001	47.4	[IDE/ SPI(t) (T.S.)*CPI]	24.77%	18.69%	6.08%
MUOS	N00039-04-C-2009	55.9	[IDE/ SPI (T.S.)]	19.23%	7.87%	11.36%
NAVSTAR GPS	FA8807-06-C-0003	86.8	[IDE/ SPI(t) (T.S.)*BEI (T.S.)]	32.89%	25.67%	7.23%
NAVSTAR GPS	FA8807-06-C-0001	87.1	[IDE/ SPI(t) (T.S.)*BEI]	33.05%	24.45%	8.60%
NAVSTAR GPS	FA8807-06-C-0004	88.1	[IDE/ SPI]	23.76%	10.33%	13.43%
WGS	FA8808-10-C-0001	96.3	IDE	29.33%	19.53%	9.79%
AEHF	F04701-02-C-0002	165.0	[IMS PD/ SPI(t)*CPI]	25.66%	23.09%	2.57%
SBIRS HIGH	F04701-95-C-0017	241.8	[IMS PD/ SPI(t) (T.S.)* BEI (T.S.)]	24.63%	21.88%	2.76%

Short Duration Contracts

Because of differences in the length of contracts it is important to analyze them separately to determine if any differences in accuracy exists. Reexamining Table 23 shows the short duration contracts (GPS OCX) and the long duration contracts (AEHF & SBIRS) have the lowest accuracy improvement (2.04%, 2.57%, 2.76%, & 3.73%). We conducted further analysis by grouping the contracts into short (GPS OCX), medium (NAVSTAR GPS, MUOS, & WGS), and long duration contracts (AEHF & SBIRS).

Table 24 shows the most accurate model is a 2.81% (23.24% vs. 20.43%) improvement over the status quo for the short duration group.

Table 24: MAPE - Short Duration Contracts (GPS OCX)

Forecasting Model	MAPE
IMS PD/SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)	20.43%
IMS PD/SPI(t)* BEI(T.S.)* CPI	20.56%
IMS PD/SPI(t) (T.S.)*BEI*CPI(T.S.)	20.56%
IMS PD/SPI*CPI*BEI(T.S.)	20.59%
IMS PD/SPI(t)*CPI*BEI	20.71%
IMS PD/SPI*CPI*BEI	20.73%
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	20.75%
IMS PD/SPI(t) (T.S.)*BEI	20.88%
IMS PD/SPI(t)(T.S.)*CPI	22.27%
IMS PD/SPI(t)(T.S.)*CPI(T.S.)	22.36%
IMS PD/SPI(t)*CPI	22.54%
IMS PD/SPI*CPI	22.56%
IMS PD/SPI(T.S.)*CPI	22.60%
IMS PD/SPI(t)*CPI(T.S.)	22.61%
IMS PD/SPI(T.S.)*CPI(T.S.)	22.64%
IMS PD/SPI(t) (T.S.)	22.66%
IMS PD/SPI(t)	22.91%
IMS PD/SPI	22.94%
IMS PD/ SPI(T.S.)	22.95%
CPR PD (status quo)	23.24%
IMS PD	23.71%
Kalman Filter	24.64%
Regression	25.04%

The range of 20.43% to 25.04% is the narrowest range of the entire analysis. Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate model. The Levene test p-value was 0.3337, denoting equal variance (see Appendix B, Figure 40). Next, we conducted a Tukey-Kramer HSD comparison of means. The connecting letters report (Figure 26) shows the most accurate model is not significantly different than the status quo. Relaxing the alpha does not separate the model from the status quo until an alpha of 0.27 (Figure 27). At this alpha level there is a much larger chance of type I error (false positive).

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.98729	0.05

Connecting Letters Report

Level		Mean
CPR PD (status quo)	A	0.23233556
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.	A	0.20425556

Levels not connected by same letter are significantly different

Figure 26: Tukey-Kramer HSD - Short Duration

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.11005	0.27

Connecting Letters Report

Level			Mean
CPR PD (status quo)	A		0.23233556
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.		B	0.20425556

Levels not connected by same letter are significantly different

Figure 27: Tukey-Kramer HSD - Short Duration (alpha = 0.27)

These results support the non-OTB group results. The two contracts analyzed do not have OTBs. What factors are affecting the accuracy improvement? Is it the length of the contract, OTBs, or a different parameter? We used regression analysis in an attempt to provide a quantitative answer to this question. The regression results are reported after the group analysis.

Medium Duration Contracts

Table 25 displays the MAPE results for medium duration contracts (six contracts). The most accurate model is an improvement of 7.8% over the status quo (27.43% vs. 19.63%). All of the forecasting models were better than the status quo except for regression. The following seven IDE models that are significantly different than the status quo:

- IDE
- IDE/ [SPI]
- IDE/ [SPI(t) (T.S.)]
- IDE/ [SPI(t) (T.S.) * BEI (T.S.)]
- IDE/ [SPI(t)]
- IDE/ [SPI(t) (T.S.) * BEI]
- IDE/ [SPI (T.S.)]

Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate models. The Levene test p-value was 0.9811, denoting equal variance (Appendix B, Figure 41). Next, we conducted a Tukey-Kramer HSD comparison of means. The seven models highlighted by the blue box in the connecting letters report (Figure 28) are significantly different from the status quo.

Referring back to Table 23, each of the six contracts in this analysis had an IDE based model as the most accurate model. When data is available, the IDE based methods appears to be the most accurate.

Table 25: MAPE - Medium Duration Contracts (MUOS, NAVSTAR GPS, & WGS)

Model	MAPE
IDE/ SPI(T.S.)	19.63%
IDE/ [SPI(t) (T.S.)*BEI]	19.68%
IDE/ SPI(t)	19.71%
IDE/ [SPI(t) (T.S.)*BEI (T.S.)]	19.77%
IDE/ [SPI(t) (T.S.)]	19.79%
IDE/ SPI	19.99%
IDE	20.95%
IMS PD/ [SPI(t) (T.S.)*BEI]	23.42%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	23.55%
IDE/ [SPI(T.S.)*CPI]	24.10%
IDE/ [SPI*CPI]	24.10%
IMS PD/ [SPI(t) (T.S.)]	24.13%
IDE/ [SPI(T.S.)*CPI(T.S.)]	24.15%
IDE/ [SPI*CPI*BEI]	24.46%
IDE/ [SPI*CPI*BEI(T.S.)]	24.48%
IMS PD/ SPI(t)	24.68%
IMS PD/ [SPI(T.S.)]	25.63%
IDE/ [SPI(t)*CPI]	26.00%
IMS PD/ SPI	26.10%
IMS PD/ [SPI*CPI*BEI]	26.21%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	26.21%
IMS PD/ [SPI*CPI*BEI(T.S.)]	26.22%
IDE/ [SPI(t)(T.S.)*CPI]	26.22%
IDE/ [SPI(t)*CPI(T.S.)]	26.26%
IMS PD	26.52%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	26.54%
IDE/ [SPI(t)*CPI*BEI]	26.57%
IDE/ [SPI(t)*CPI*BEI(T.S.)]	26.59%
IDE/ [SPI(t)(T.S.)*CPI(T.S.)]	26.59%
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	26.60%
IMS PD/ [SPI(t)*CPI]	26.66%
Kalman Filter	26.67%
IMS PD/ [SPI(t)*CPI*BEI]	26.72%
IMS PD/ [SPI(t)*CPI(T.S.)]	26.78%
IMS PD/ [SPI(t)(T.S.)*CPI]	26.93%
IMS PD/ [SPI*CPI]	26.93%
IMS PD/ [SPI(T.S.)*CPI]	26.94%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	27.01%
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)]	27.01%
IDE/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	27.22%
IDE/ [SPI(t) (T.S.)*BEI(T.S.)*CPI(T.S.)]	27.25%
CPR PD (status quo)	27.43%
Regression	39.81%

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
3.16547	0.05

Connecting Letters Report

Level			Mean
CPR PD (Staus Quo)	A		0.27428963
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	A	B	0.23554272
IMS PD/SPI(t) (T.S.)*BEI	A	B	0.23423630
Independent Duration Estimate (IDE		B	0.20954840
IDE/SPI		B	0.19986593
IDE/SPI(t) (T.S.)		B	0.19789852
IDE/SPI(t) (T.S.)*BEI (T.S.)		B	0.19768173
IDE/SPI(t)		B	0.19706667
IDE/SPI(t) (T.S.)*BEI		B	0.19678543
IDE/ SPI(T.S.)		B	0.19626963

Levels not connected by same letter are significantly different

Figure 28: Tukey-Kramer HSD - Medium Duration

Why are the results different for short and medium duration contracts? The medium duration models are more likely to have an OTB than the shorter contracts (4 of 6 compared to zero). Contracts with OTBs appear to have less accurate status quo estimates compared to non OTB contracts. The effect of OTBs are explored further in the regression analysis section. Regardless of the reason, there is clear evidence that any of the seven IDE based models are the most accurate models for medium duration contracts (47.4 to 96.3 months).

Long Duration Contracts

Table 26 displays the accuracy results for long duration contracts (AEHF and SBIRS). The results were less substantial for the longer contracts with only a 2.13% (25.05% vs. 22.92%) improvement over the status quo.

Table 26: MAPE - Long Duration Contracts (AEHF & SBIRS)

Metric	MAPE
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	22.92%
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)]	22.94%
IMS PD/ [SPI*CPI*BEI(T.S.)]	22.98%
IMS PD/ [SPI(t)*CPI*BEI]	23.01%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	23.04%
IMS PD/ [SPI*CPI*BEI]	23.08%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	23.18%
IMS PD/ [SPI(t) (T.S.)*BEI]	23.28%
IMS PD/ [SPI(t)*CPI(T.S.)]	23.58%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	23.59%
IMS PD/ [SPI(t)*CPI]	23.59%
IMS PD/ [SPI(t)(T.S.)*CPI]	23.59%
IMS PD/ [SPI(T.S.)*CPI]	23.68%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	23.68%
IMS PD/ [SPI*CPI]	23.69%
IMS PD/ SPI(t)	24.69%
IMS PD/ [SPI(t) (T.S.)]	24.70%
IMS PD/ SPI(T.S.)	24.78%
IMS PD/ SPI	24.78%
CPR PD (status quo)	25.05%
IMS PD	25.17%
Kalman Filter	25.27%
Regression	34.03%

Analyzing all of the models at once resulted in unequal variances. We truncated the analysis to include the CPR PD and the most accurate model. The Levene test p-value was 0.0714, denoting equal variance (Appendix B, Figure 42). Next, we conducted a Tukey-Kramer HSD comparison of means. The connecting letters report (Figure 29) shows the most accurate model is not significantly different than the status quo. After relaxing the alpha (0.10) the most accurate model (IMS PD / [SPI(t)*CPI *BEI (T.S.)) is significantly different than the status quo (Figure 30).

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.96331	0.05

Connecting Letters Report

Level			Mean
CPR PD (Status Quo)	A		0.25047893
IMS PD/SPI(t)*CPI*BEI(T.S.	A		0.22921011

Levels not connected by same letter are significantly different

Figure 29: Tukey-Kramer HSD - Long Duration

Means Comparisons

Comparisons for all pairs using Tukey-Kramer HSD

Confidence Quantile

q*	Alpha
1.64700	0.1

Connecting Letters Report

Level			Mean
CPR PD (Status Quo)	A		0.25047893
IMS PD/SPI(t)*CPI*BEI(T.S.		B	0.22921011

Levels not connected by same letter are significantly different

Figure 30: Tukey-Kramer HSD - Long Duration (Alpha = 0.10)

Once again, a larger alpha means there is a greater chance of type I error. Therefore, the difference between the most accurate model and the status quo is not as pronounced as the prior analysis with a smaller alpha. Why is this the case? The potential reasons are the most confounding of this analysis. For SBIRS, BEI data were not available until 89 months into the contract (37% complete). BEI based models have been among the best performers and each of the six models most accurate models here contain a BEI parameter. SBIRS was the only contract out of seven (with IDE data) that did not have an IDE based model as the most accurate. One reason may have been data availability; IDE data was not available until 141 months (58% complete). For AEHF, IDE was not available at all. Furthermore, the first EVM data were not reported until 11 months into the contract. Another possibility, the schedule and cost performance factors are not the drivers of schedule growth. Perhaps management decisions are driving the cost and schedule growth of the longer contracts. Therefore enhancing the IMS PD with performance factors will not drastically improve the estimate accuracy. Another factor to consider, these program are not 100% complete. The forecast accuracy results will be different if the actual completion date is different than the current planned completion dates for AEHF and SBIRS: 06/30/2015 and 12/31/2016.

Sensitivity Analysis: Entire Data Set

In the next section what-if analysis is conducted because there is no single dominant forecasting model. The scenario is what if we use the most accurate overall model then examine how well it fares compared to the status quo for each contract. Table 27 displays the what-if analysis for the most accurate IMS based model [(IMS PD / (SPI(t) (T.S.) * BEI)] applied to all contracts (refer to Table 19 for the most accurate IMS models).

Nine out of the ten contracts show an improvement in accuracy over the status quo. The WGS contract (FA8808-10-C-0001) is the only contract in the entire analysis where an IMS index based does not improve upon the status quo. This contract had high CPI, SPI, and SPI(t) early in the contract (see Table 28). This resulted in the models predicting the contract would be completed faster than the planned duration. This large error could not be overcome by improved accuracy in the later periods.

Table 27: Comparison of Status Quo vs. Most Accurate IMS Model

Program	Contract	CPR PD (status quo)	[IMS PD / SPI(t) (T.S.) *BEI]	Delta	Signif. difference
Non-OTB Contracts					
GPS OCX	FA8807-08-C-0001	20.41%	18.87%	1.54%	No
GPS OCX	FA8807-08-C-0003	25.71%	22.65%	3.06%	No
WGS	FA8808-06-C-0001	24.77%	22.86%	1.91%	No
WGS	FA8808-10-C-0001	29.33%	30.90%	-1.57%	No
OTB Contracts					
AEHF	F04701-02-C-0002	25.66%	25.11%	0.55%	No
MUOS	N00039-04-C-2009	19.23%	14.22%	5.01%	Yes
NAVSTAR GPS	FA8807-06-C-0001	33.05%	30.52%	2.53%	No
NAVSTAR GPS	FA8807-06-C-0003	32.89%	29.21%	3.69%	No
NAVSTAR GPS	FA8807-06-C-0004	23.76%	14.92%	8.84%	Yes
SBIRS	F04701-95-C-0017	24.63%	22.03%	2.60%	No

Table 28: WGS (FA8808-10-C-0001) – Index Values

Month Count	CPI	CPI (T.S.)	SPI	SPI (T.S.)	SPI(t)	SPI(t) (T.S.)
1	1.215	1.215	1.166	1.166	1.204	1.204
2	1.231	1.223	1.328	1.247	1.417	1.311
3	1.231	1.226	1.475	1.629	1.289	1.303
4	1.195	1.218	1.351	1.330	1.266	1.324
5	1.165	1.170	1.302	1.325	1.217	1.279
6	1.124	1.108	1.256	1.347	1.249	1.274
7	1.112	1.121	1.307	1.393	1.234	1.268
8	1.095	1.113	1.284	1.343	1.197	1.259
9	1.081	1.083	1.255	1.350	1.188	1.251
10	1.081	1.097	1.207	1.264	1.149	1.241
11	1.077	1.094	1.208	1.262	1.149	1.129
12	1.059	1.051	1.182	1.225	1.147	1.100

Because of the similarities in accuracy the IMS model was only significantly different than the status quo in two out of ten contracts. Despite less than overwhelming results, the best IMS based model [(IMS PD / (SPI(t) (T.S.) * BEI)] is no worse in eight of ten contracts and more accurate in two of ten contracts.

Sensitivity Analysis: IDE Data Set

Table 29 displays another what-if scenario. This time, we choose to use one of the most accurate overall models [IDE/ SPI(t)] for contracts with IDE data. The result is an improvement in all seven contracts.

Table 29: Comparison of Status Quo vs. Most Accurate Model with IDE Data

Program	Contract	CPR PD (status quo)	IDE/ SPI(t)	Delta	Signif. Diff.
		Non-OTB Contracts			
WGS	FA8808-06-C-0001	24.77%	20.05%	4.72%	Yes
WGS	FA8808-10-C-0001	29.33%	21.65%	7.68%	Yes
		OTB Contracts			
MUOS	N00039-04-C-2009	19.23%	8.29%	10.94%	Yes
NAVSTAR GPS	FA8807-06-C-0001	33.05%	25.98%	7.07%	Yes
NAVSTAR GPS	FA8807-06-C-0003	32.89%	26.71%	6.18%	No
NAVSTAR GPS	FA8807-06-C-0004	23.76%	13.25%	10.51%	Yes
SBIRS	F04701-95-C-0017	24.63%	24.49%	0.14%	No

A very small improvement was achieved in SBIRS (0.14%). However, a more substantial improvement (4.72% to 10.94%) was achieved for the other contracts. Five of the seven contracts have improved accuracy and the model is significantly different than the status quo. Obviously, SBIRS was not significantly different (0.14% difference). The primary reason was previously discussed (IDE data not available until 58% complete). At first glance, NAVSTAR GPS (FA8807-06-C-0003) was expected to be significantly different (6.18%). Upon closer inspection, IDE data were not available until 18 months into the contract (26% complete). Another IDE data lapse occurred from month 52 to 71. Thus, the status quo and IDE / SPI(t) are more similar for this contract than the accuracy results would suggest.

Regression Analysis

The results of the preceding analysis exhibited differences between the accuracy of the CPR PD (status quo), IMS models, and IDE models. Why do these differences occur? How does the length of the contract, OTBs, budget size, cost growth, and schedule growth affect duration estimate accuracy? We used regression analysis in an attempt to provide quantitative answers to these questions. First, we divided the dataset into the following dependent variables and data sets:

- CPR PD Accuracy (All Contracts)
- Most Accurate Model for Each Contract (All Contracts)
- Most Accurate Model for the Seven Contract with IDE Data (7 of 10 contracts)
- IMS Delta Compared to CPR PD (All Contracts)
- IDE Delta Compared to CPR PD (7 of 10 contracts)
- IDE Delta Compared to IMS (7 of 10 contracts)

Table 30 lists the data set for this analysis. Table 31 through Table 36 summarize the results of the regression analysis. Each of the models met the following diagnostics:

- Studentized residuals check for outliers (no observations greater than 3 standard deviations)
- Cook’s D influence (less than 0.5)
- Shapiro-Wilk test for Normality of Residuals (p-value greater than 0.05)
- Breusch-Pagan test for heteroscedasticity (p-value greater than 0.05)

The supporting documentation for the regression analysis and diagnostics is located in Appendix D, beginning with Figure 53.

Table 30: Regression Analysis Data Set

Program	Contract	Initial BAC (BY97$)	Final BAC (BY97$)	OTBs	CPR PD (%)	IMS (%)	IDE (%)
AEHF	F04701-02-C-0002	2395.9M	5481.6M	3	25.7	23.1	N/A
GPS OCX	FA8807-08-C-0001	119.0M	142.5M	0	20.4	18.4	N/A
GPS OCX	FA8807-08-C-0003	118.6M	141.0M	0	25.7	22.0	N/A
MUOS	N00039-04-C-2009	70.3M	77.1M	3	19.2	14.0	7.9
Navstar GPS	FA8807-06-C-0001	20.8M	94.0M	1	33.1	25.1	24.5
Navstar GPS	FA8807-06-C-0003	29.8M	79.5M	4	32.9	26.1	25.7
Navstar GPS	FA8807-06-C-0004	47.8M	86.9M	1	23.8	14.9	10.3
SBIRS	F04701-95-C-0017	1663.6M	6383.6M	4	24.6	21.9	24.4
WGS	FA8808-10-C-0001	115.2M	120.6M	0	29.3	29.5	19.5
WGS	FA8808-06-C-0001	295.8M	734.3M	0	24.8	20.3	18.7

Regression Analysis: CPR PD (status quo) Accuracy

Table 31 shows the regression results for the accuracy of the CPR PD (status quo). The accuracy of the status quo estimate was correlated with the reciprocal of schedule growth (1/schedule growth). This transformation is non-linear, as schedule growth increases the CPR PD accuracy decreases at a diminishing rate (Figure 31). To reiterate a discussion from chapter one, the largest sources of schedule growth are estimating errors or decisions affecting the schedule. For these ten contracts, schedule growth may occur if the initial estimates are overly optimistic and/or decisions are made that affect the schedule. In theory, greater schedule growth (regardless of the reason) leads to less schedule data fidelity resulting in less accurate status quo schedule estimates.

Table 31: CPR PD (status quo) Accuracy

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
1/Sched Growth	0.5835	0.0061	-3.7	-0.79	Yes (0.25)	0.4501	0.7794	8.3%	7.9%

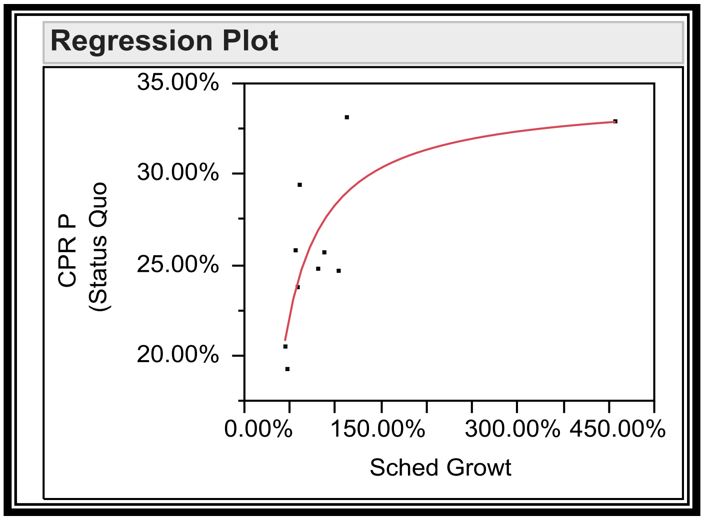


Figure 31: CPR PD

Regression Analysis: IMS and IDE Accuracy

Table 32 and Table 33 list the regression results for the accuracy of the IMS models (all contracts) and IDE data set respectively.

Table 32: Most Accurate Models - All Contracts

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
OTB & Sched Growth DV	0.5391	0.0094	-3.4	-0.77	Yes (0.29)	0.9402	0.2756	10.1%	6.9%

The most significant parameter for both data sets was the combination of at least one OTB and low schedule growth (less than 62%) into one indicator variable (two contracts in this cohort). The two contracts satisfying both of these conditions experienced increased accuracy gains. As previously discussed, high schedule growth may be the result of estimate errors and decisions. Consequently, lower schedule growth may indicate better initial estimates and management decisions are playing a lesser role. Therefore, the data may better explain the contract's performance leading to more accurate schedule estimates (IMS and IDE models).

Table 33: Most Accurate Model – IDE Data Set

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
OTB & Sched Growth DV	0.8260	0.0029	-5.4	-0.92	Yes (0.27)	0.3235	0.3696	13.0%	12.2%

On the other hand, OTBs may confound our results. Our earlier analysis showed contracts with an OTB exhibited improved accuracy compared to contracts without an OTB. That result is also supported by the regression analysis. Due to the complexity of MDAPs and our limited data set it is difficult to tease out simple explanations. OTB research from 2010 concluded contracts undergoing an OTB did not improve cost performance (Jack, 2010). However, cost estimate research from 2009 found increased accuracy for estimating the EAC of OTB contracts (Trahan, 2009). The regression results from our analysis suggest some contracts that undergo an OTB may gain fidelity in EVM schedule indices and the integrated master schedule (IMS). This potential fidelity may be detected by the IMS PD and IDE models, but not the status quo. If this is true, the models researched here may be more useful for OTB contracts. Further research is necessary to provide a more definitive answer.

Regression Analysis: IMS Accuracy Delta

Examining the difference between the IMS models and the CPR PD yields slightly different results (Table 34). If a program has just one OTB there is an increase in accuracy (IMS over status quo). This result may support the hypothesis that having one OTB is beneficial, but undergoing additional OTBs does not improve schedule performance.

Table 34: IMS Delta (All Contracts)

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
1 OTB DV	0.4956	0.0139	3.1	0.74	Yes (0.26)	0.9705	0.2974	280%	28.8%

Regression Analysis: IDE Accuracy Delta

The summary regression results for the IDE delta data set are listed in Table 35. Having schedule growth under 62% and one OTB was significant. Both variables increased the accuracy delta (IDE compared to status quo). The 1 OTB dummy variable by itself was no longer significant and the schedule growth dummy variable had a stronger impact than the 1 OTB DV. The schedule growth dummy variable (under 62%) by itself was significant (three contracts). Once again, high schedule growth may be the result of estimating errors and/or decisions. Therefore, lower schedule growth may indicate the opposite, leading to better data fidelity. A more thorough explanation is beyond the scope of this research, further research is necessary to explore this relationship. Whatever the reasons, the accuracy improvement (over status quo) is more pronounced for contracts with low schedule growth and one OTB.

Table 35: IDE Delta (7 of 10 contracts)

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
• **Sched Growth DV**	0.7574	0.0262	3.8	0.77	Yes (0.46)	0.6839	0.7690	21.4%	7.5%
• **1 OTB DV**			2.1	0.42					
• **Sched Growth DV**	0.5923	0.0263	3.1	0.81	No (0.51)	0.7642	0.5318	28.8%	14.7%

Regression Analysis: IDE – IMS Accuracy Delta

The summary regression results for the IDE - IMS data set are listed in Table 36. The difference between IDE and IMS accuracy is the greatest when cost growth is low. The larger the natural log of a contract's cost growth, the lower the increase in accuracy (IDE - IMS). Because it's a natural log transformed parameter, the effect diminishes as the cost growth increases (see Figure 32 for a visual depiction).

Table 36: IDE – IMS Accuracy Delta

Term(s)	Adj R^2	Model p-value	t ratio	Std Beta	Cook's D (< 0.5)	Shapiro Wilk p-value	Breusch-Pagan p-value	MAPE	Median APE
• **Log (Cost Growth)**	0.923	0.0027	-8.6	-1.04	Yes (0.43)	0.5255	0.4964	37.0%	20.4%
• **1 OTB DV**			2.6	0.31					
• **Log (Cost Growth)**	0.838	0.0024	-5.7	-0.93	Yes (0.45)	0.8707	0.6268	72.0%	42.0%

Why do larger cost growth contracts exhibit a smaller advantage for the IDE models (over the IMS models)? One possible explanation is cost growth is similar to schedule growth; if large cost growth occurs, management decisions may be playing a larger role in explaining the schedule than the contract's data. Contracts with high cost growth may lose schedule data fidelity; therefore the IDE models lose their accuracy advantage over the IMS models. On the other hand, low cost growth may have the opposite effect and better data fidelity. Once again, contracts with only one OTB have a slight accuracy gain. This result may further support the hypothesis that having one OTB is beneficial, but more than one is not. Additional research is necessary to explore the relationship between OTBs, cost growth, and schedule estimate accuracy.

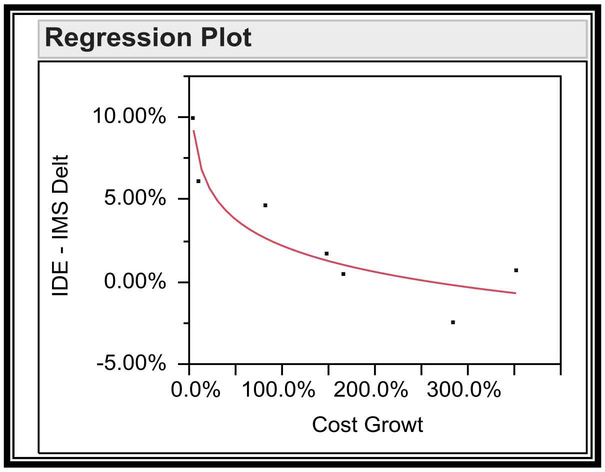


Figure 32: IDE - IMS Accuracy Delta

Regression Analysis Summary

In summary, OTBs, schedule growth, and cost growth were the dominant variables explaining the accuracy of the duration estimating models (listed in Table 37).

Table 37: Variables Effect on Accuracy

Response	Improves Accuracy	Reduces Accuracy
CPR PD Accuracy	Low schedule growth	Increasing schedule growth reduces accuracy at a diminishing rate (non-linear)
IMS Accuracy	Contracts with an OTB and schedule growth under 62%	Contracts with OTB ≠ 1 and schedule growth over 62%
IDE Accuracy	Contracts with an OTB and schedule growth under 62%	Contracts with OTB ≠ 1 and schedule growth over 62%
IMS – CPR PD Delta	Contracts with OTB =1	Contracts with OTB ≠ 1
IDE – CPR PD Delta	Contracts with schedule growth under 62% and OTB =1	Contracts with schedule growth over 62% and OTB ≠ 1
IDE – IMS Delta	Low cost growth	Increasing cost growth reduces accuracy at a diminishing rate (non-linear)

Schedule growth is correlated with less accuracy for the CPR PD (status quo). The accuracy of the IMS PD and IDE models is correlated with OTBs and schedule growth. The accuracy improvement between the IMS models and the CPR PD is largest for contracts with one OTB. The accuracy improvement between IDE and the CPR is largest for contracts with one OTB and low schedule growth (less than 62%). Finally, the accuracy improvement between IDE and IMs is greatest for low cost growth contracts. It should be noted there are substantial limitations with the regression results; the sample size is small and there are many possible explanations for the differences in the accuracy delta besides the variables examined here. We cannot conclude that OTBs, schedule growth, and cost growth directly impact the duration estimate accuracy, but they are correlated for our data set. The relationships are discussed here to provide a quantitative explanation for differences in the accuracy of the duration estimates and may serve as a guide to help practitioners decide when to use each model.

Forecast Model Timeliness

The next section discusses the timeliness of the IMS forecasts. Table 38 displays the MAPE over time intervals (from 0% to 100%). Table 38 is highlighted with a heat map: dark green is favorable (10th percentile), yellow is average (50th percentile), and dark red is unfavorable (90th percentile). The more dark green present, the more accurate the model. Each of the models exhibit improved accuracy as the contract matures. Early in a contract there is more uncertainty, therefore the early estimates are inherently less accurate than later estimates. The status quo is one of the least accurate methods (red) from 0% to 70%. The lack of accuracy for status quo estimates may be the result of early estimating errors or management decisions as previously discussed. From 0 to 60% complete the IMS/ (SPI(t)*CPI*BEI) based metrics are the most accurate (including time series variations). These models are the most pessimistic because they contain three performance factors; most of the contracts experienced less than favorable cost and schedule performance (index values less than one). With the exception of the WGS Block 2 Follow On contract, the contracts in this analysis did not have favorable metrics in the early periods. Therefore, the pessimistic duration estimates were higher than the status quo. The accuracy of pessimistic models should be no surprise considering every contract experienced schedule growth. The pessimistic models incorporate performance factors and detect schedule growth earlier than the status quo method. Therefore, using a pessimistic forecast model in the early periods (0 to 60%) should improve the accuracy of duration estimates.

From 61% to 70% the most accurate models are: IMS PD/ (SPI(t)*BEI) and IMS PD/ (SPI*CPI*BEI) (including time series). These models are less pessimistic, but still incorporate cost and schedule performance into the model. The difference between the most accurate model from 0% to 60% [IMS/ (SPI(t)*CPI*BEI)] and 61% to 70% [IMS PD/ (SPI(t)*BEI)] is the removal of the CPI. The other model [IMS PD/ (SPI*CPI*BEI)] replaces SPI(t) with SPI and is therefore a less pessimistic model because SPI begins to converge to 1 as the program matures. As a contract matures the (relatively) less pessimistic models become more accurate. From 71% to 100% complete the following models are the most accurate: IMS PD, IMS PD/ SPI(t) (including time series) and IMS PD/ SPI (including time series). At this point in the contract the performance factors (SPI and SPI(t)) were close to one (contract is on schedule); therefore they are not improving the accuracy of the basic IMS PD. On the other end of the spectrum, the most accurate models from 0% to 60% are now the worst performers.

Table 38: MAPE at Time Intervals (All Contracts)

Forecasting Model	Percent Complete Interval									
	0 to 10	11 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 70	71 to 80	81 to 90	91 to 100
CPR PD (status quo)	49.2%	48.1%	39.4%	34.2%	29.9%	26.1%	17.7%	10.1%	5.0%	4.4%
IMS PD	49.0%	47.8%	38.7%	34.0%	30.3%	26.5%	17.8%	10.5%	3.4%	1.3%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	48.2%	46.6%	37.5%	31.2%	25.5%	18.9%	12.3%	9.7%	6.0%	2.2%
IMS PD/ [SPI(t) (T.S.)*BEI(T.S.)*CPI(T.S.)]	47.9%	45.7%	37.0%	**28.7%**	**22.3%**	16.1%	15.2%	17.5%	7.9%	11.3%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	47.9%	**45.6%**	37.0%	28.7%	22.3%	16.2%	15.2%	17.5%	8.0%	11.3%
IMS PD/ [SPI(t) (T.S.)*BEI]	48.2%	46.5%	37.5%	31.3%	25.6%	18.9%	12.3%	9.7%	6.1%	2.2%
IMS PD/ [SPI(t) (T.S.)]	48.4%	47.0%	37.8%	33.0%	29.1%	22.9%	14.7%	**8.8%**	**3.2%**	2.0%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	48.1%	46.1%	36.6%	30.7%	26.7%	19.6%	13.0%	12.2%	5.1%	10.9%
IMS PD/ [SPI(t)(T.S.)*CPI]	48.2%	46.2%	**36.4%**	30.5%	26.1%	19.7%	16.7%	14.0%	4.9%	10.9%
IMS PD/ [SPI(t)*CPI(T.S.)]	47.6%	46.4%	37.4%	30.6%	26.2%	19.1%	15.5%	14.2%	5.0%	11.3%
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	**47.6%**	46.2%	36.9%	29.1%	22.7%	**15.8%**	14.9%	16.9%	7.9%	11.5%
IMS PD/ [SPI(t)*CPI*BEI]	48.0%	45.7%	36.8%	29.0%	22.6%	16.0%	15.0%	17.5%	8.0%	11.6%
IMS PD/ [SPI(t)*CPI]	47.7%	46.5%	37.2%	30.9%	26.3%	18.9%	15.1%	14.2%	4.8%	11.2%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	48.0%	46.8%	37.1%	30.5%	26.6%	21.0%	13.9%	13.1%	5.0%	10.1%
IMS PD/ [SPI(T.S.)*CPI]	48.1%	46.8%	37.1%	30.8%	26.7%	21.0%	14.1%	13.0%	4.9%	9.9%
IMS PD/ [SPI*CPI*BEI(T.S.)]	47.8%	46.4%	36.8%	29.2%	23.1%	17.2%	12.7%	15.7%	7.8%	10.1%
IMS PD/ [SPI*CPI*BEI]	47.8%	46.3%	36.8%	29.2%	23.1%	17.2%	12.7%	15.7%	7.8%	10.1%
IMS PD/ [SPI*CPI]	48.0%	46.8%	37.1%	31.0%	26.7%	21.0%	13.9%	13.1%	4.7%	9.8%
IMS PD/ SPI	48.1%	47.6%	37.8%	33.3%	29.8%	26.0%	17.4%	10.2%	3.3%	**1.2%**
IMS PD/ SPI(t)	47.9%	47.3%	37.9%	33.2%	29.4%	23.3%	15.0%	9.0%	3.2%	2.2%
IMS PD/ SPI(T.S.)	48.3%	47.7%	37.7%	33.0%	29.8%	25.9%	17.3%	10.1%	3.4%	1.3%
Kalman	48.9%	48.0%	38.6%	34.1%	30.0%	25.1%	18.7%	10.3%	5.2%	2.7%
Regression	52.4%	54.2%	49.7%	44.0%	41.9%	38.6%	33.2%	21.5%	15.9%	13.2%

The next section discusses the timeliness of the IDE and IMS forecasts for the seven contracts with IDE data. Table 39 and Table 40 display the MAPE over time intervals (from 0% to 100%). There is not a single dominant model across the all intervals. This discussion should provide insight into which models perform best at certain intervals.

Table 39: MAPE at Time Intervals (with IDE Data)

Forecasting Model	Percent Complete Interval									
	0 to 10	11 to 20	21 to 30	31 to 40	41 to 50	51 to 60	61 to 70	71 to 80	81 to 90	91 to 100
CPR PD (status quo)	51.1%	49.7%	39.7%	35.0%	29.5%	26.3%	17.6%	9.0%	4.7%	5.2%
IMS PD	50.7%	49.2%	38.7%	34.8%	30.1%	26.6%	17.8%	9.3%	2.5%	1.0%
IMS PD/ SPI(T.S.)	49.9%	49.2%	37.8%	33.7%	29.6%	26.1%	17.1%	9.0%	2.5%	1.0%
IMS PD/SPI	49.7%	49.2%	37.9%	34.1%	29.6%	26.0%	17.2%	9.1%	2.4%	**0.9%**
IMS PD/SPI(t)	49.5%	48.9%	38.1%	34.0%	29.2%	22.6%	14.1%	7.6%	2.4%	2.2%
IMS PD/ [SPI(t) (T.S.)*BEI]	50.0%	48.2%	38.1%	31.6%	24.2%	17.0%	10.7%	8.5%	6.1%	2.3%
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.)]	50.0%	48.2%	38.1%	31.3%	24.3%	17.1%	10.6%	8.1%	6.3%	2.4%
IMS PD/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	49.6%	47.3%	38.5%	30.0%	22.1%	14.5%	13.9%	18.4%	8.5%	13.6%
IMS PD/SPI(t) T.S.	50.0%	48.4%	38.1%	33.8%	28.9%	22.2%	13.7%	7.3%	**2.3%**	1.9%
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.)]	49.6%	47.3%	38.4%	29.7%	22.2%	14.6%	13.8%	17.9%	8.6%	13.7%
IMS PD/ [SPI(t)(T.S.)*CPI(T.S.)]	49.6%	47.5%	37.6%	32.3%	27.8%	18.9%	11.0%	11.2%	4.6%	12.9%
IMS PD/ [SPI(t)(T.S.)*CPI]	49.8%	47.7%	37.4%	32.1%	26.9%	19.1%	15.8%	13.6%	4.3%	13.0%
IMS PD/ [SPI(t)*CPI]	49.3%	48.2%	38.3%	32.6%	27.2%	18.0%	13.8%	13.9%	4.4%	13.3%
IMS PD/ [SPI(t)*CPI(T.S.)]	**49.1%**	48.1%	38.4%	32.3%	27.1%	18.3%	14.2%	14.0%	4.6%	13.5%
IMS PD/ [SPI(t)*CPI*BEI]	49.7%	47.6%	38.2%	30.3%	22.4%	14.3%	13.7%	18.4%	8.5%	14.0%
IMS PD/ [SPI(t)*CPI*BEI(T.S.)]	49.3%	48.1%	38.2%	30.1%	22.6%	14.2%	13.4%	17.2%	8.6%	14.0%
IMS PD/ [SPI(T.S.)*CPI(T.S.)]	49.5%	48.3%	38.1%	32.0%	27.6%	20.8%	12.2%	12.2%	4.5%	12.0%
IMS PD/ [SPI(T.S.)*CPI]	49.7%	48.4%	38.0%	32.4%	27.6%	20.7%	12.4%	12.2%	4.3%	11.7%
IMS PD/ [SPI*CPI]	49.5%	48.5%	38.0%	32.7%	27.6%	20.8%	12.2%	12.2%	4.1%	11.5%
IMS PD/ [SPI*CPI*BEI]	49.5%	48.3%	38.0%	30.5%	23.0%	15.8%	10.7%	15.8%	8.2%	12.0%
IMS PD/ [SPI*CPI*BEI(T.S.)]	49.5%	48.4%	38.0%	30.3%	23.0%	15.9%	10.6%	15.3%	8.3%	12.1%

From 0% to 60% the status quo is among the least accurate. The 0% to 10% interval MAPEs are close across the board with the IMS PD based metrics having a slight edge. From 11% to 60% completion the following models are the most accurate:

- IDE/ (SPI(t)*BEI) (including time series)
- IDE/ (SPI*CPI*BEI) (including time series)

This result is similar to the prior section's analysis. However, these performance factors are less pessimistic than the SPI(t)*CPI*BEI. The IDE by itself is a pessimistic model because it modifies the IMS PD by adding the schedule slip. Applying a moderately pessimistic performance factor to the IDE will further improve the forecast accuracy.

From 61% to 80% complete the following models are the most accurate: IDE, IDE/ SPI(t), and IDE/ SPI. Again, the results are similar to the prior section's analysis, less pessimistic performance factors become more accurate as the contract matures. From 81% to 100% complete the following models are the most accurate: IMS PD, IMS PD/ SPI(t), and IMS PD/ SPI. Once again, as the program matures to the later stages the basic forecast is among the most accurate. Over the same time interval many of the IDE based metrics lose their accuracy advantage because they are overestimating duration.

Table 40: MAPE at Time Intervals (with IDE Data)

	Percent Complete Interval									
Forecasting Model	**0 to 10**	**11 to 20**	**21 to 30**	**31 to 40**	**41 to 50**	**51 to 60**	**61 to 70**	**71 to 80**	**81 to 90**	**91 to 100**
IDE	50.7%	48.2%	35.0%	25.0%	23.5%	16.0%	**8.9%**	**5.3%**	6.5%	2.7%
IDE/SPI(T.S.)	49.9%	48.2%	34.4%	25.1%	23.0%	15.5%	**8.9%**	5.4%	6.5%	2.7%
IDE/SPI	51.3%	48.4%	**33.7%**	25.2%	23.0%	15.4%	**8.9%**	5.4%	6.4%	2.7%
IDE/SPI(t)	50.9%	48.1%	33.9%	25.2%	22.6%	12.9%	10.0%	7.5%	6.3%	4.0%
IDE/ [SPI(t) (T.S.)*BEI (T.S.)]	51.0%	47.8%	34.1%	24.0%	17.5%	**10.5%**	15.1%	12.4%	10.1%	4.0%
IDE/SPI(t) (T.S.)	51.0%	48.0%	34.1%	25.3%	22.3%	12.7%	9.8%	7.8%	5.9%	3.6%
IDE/ [SPI(t) (T.S.)*BEI]	51.0%	47.4%	34.0%	24.5%	17.2%	10.8%	15.5%	12.7%	9.7%	3.9%
IDE/ [SPI(t) (T.S.)*BEI(T.S.)*CPI(T.S.)]	51.7%	47.6%	34.8%	24.3%	17.3%	16.1%	25.3%	25.7%	15.8%	13.8%
IDE/ [SPI(t) (T.S.)*BEI*CPI(T.S.)]	51.7%	**47.2%**	34.7%	24.9%	**17.0%**	16.4%	25.8%	26.0%	15.1%	13.7%
IDE/ [SPI(t)(T.S.)*CPI]	51.6%	47.7%	34.8%	25.2%	21.4%	16.3%	19.0%	20.4%	11.2%	13.5%
IDE/ [SPI(t)(T.S.)*CPI(T.S.)]	51.7%	47.8%	34.8%	25.5%	21.4%	16.8%	19.4%	20.6%	11.1%	13.4%
IDE/ [SPI(t)*CPI]	51.5%	47.9%	34.6%	25.2%	21.6%	16.4%	18.6%	19.9%	11.3%	14.0%
IDE/ [SPI(t)*CPI(T.S.)]	51.6%	48.0%	34.7%	25.4%	21.7%	16.9%	19.0%	20.0%	11.2%	13.9%
IDE/ [SPI(t)*CPI*BEI]	51.5%	47.3%	34.5%	24.1%	**17.0%**	16.0%	25.0%	25.3%	15.3%	14.3%
IDE/ [SPI(t)*CPI*BEI(T.S.)]	51.5%	47.7%	34.7%	**23.6%**	17.2%	15.8%	24.5%	25.0%	16.0%	14.4%
IDE/ [SPI(T.S.)*CPI]	50.5%	47.9%	35.2%	25.2%	21.9%	12.6%	11.6%	17.2%	12.4%	12.4%
IDE/ [SPI(T.S.)*CPI(T.S.)]	50.6%	48.0%	35.2%	25.5%	21.9%	13.1%	11.8%	17.3%	12.3%	12.3%
IDE/ [SPI*CPI]	51.9%	48.2%	34.4%	25.2%	21.9%	12.6%	11.6%	17.0%	12.3%	12.4%
IDE/ [SPI*CPI*BEI]	51.9%	47.6%	34.3%	24.1%	**17.0%**	11.4%	17.6%	22.4%	16.3%	12.7%
IDE/ [SPI*CPI*BEI(T.S.)]	51.9%	48.1%	34.4%	**23.6%**	17.3%	11.1%	17.2%	22.0%	17.0%	12.8%
Kalman Filter	50.8%	49.6%	38.8%	35.1%	29.9%	24.5%	17.3%	9.0%	4.0%	2.8%
Regression	54.7%	55.3%	54.9%	44.4%	41.4%	40.9%	35.7%	23.9%	18.3%	14.8%

Validating the Cost Estimating Model

The final area of analysis applies a duration estimate to the BCWP burn rate model in order to assess the cost estimate accuracy. This section is ancillary to the primary research, but is related to the overall research objective. The genesis of this research was to improve the accuracy of the BCWP based cost estimate by improving the accuracy of the duration estimate. Due to time constraints only one of the duration models was tested. This model [IMS PD / (SPI(t)*CPI)] was selected due its simplicity and relative ease of calculation. The five contracts listed in Table 41 were added to the original database to validate the cost model.

Table 41: Additional Contracts for Cost Model Validation

Program	Contract	Type
FAB-T (Family of Beyond Line-of-Sight Terminals)	F19628-02-C-0048	RDT&E
MUOS (Mobile User Objective System)	N00039-04-C-2009	RDT&E
GPS OCX (Next Generation Control Segment)	FA8807-10-C-0001	RDT&E
MGUE (Military GPS User Equipment)	FA8807-12-C-0011	RDT&E
EELV (Evolved Expendable Launch Vehicle)	FA8811-13-C-0001	Production

FAB-T is a completed contract and met the initial screening parameters, but it was reported as 61% complete therefore it was not included in the schedule database. MUOS was not readily available via DCARC, but was obtained from the author of the AFCAA study (Keaton, 2014). The data were obtained too late in the research process to be included in the schedule database; however, the data could be included in the cost estimate validation. MGUE and GPS OCX Phase B were eliminated in the original schedule data filter because they were not complete or near complete (at least 90%). These contracts were included in the cost estimate validation to test the model on less mature contracts (less than 90% complete). Finally, EELV was selected to test the model on a completed production contract.

Three cost estimating models were analyzed:

- Reported EAC: [Contractor reported EAC]
- BCWP1: [CPR PD and Actual Time] (Keaton, 2014)
- BCWP2: [IMS PD / (SPI(t)*CPI) and actual time]

The Reported EAC is a base case for comparison purposes. BCWP1 is the model from the AFCAA research (Keaton, 2014). BCWP2 uses the same BCWP burn rate and actual BCWP to date as BCWP1. However, BCWP2 applies a duration model estimate from this research. The cost estimate MAPE is calculated as follows:

Equation 46: Final EAC MAPE

$$\mathrm{MAPE} = \frac{(\mathrm{EAC_{Final}} - \mathrm{EAC_{Forecast}})}{\mathrm{EAC_{Final}}} * 100$$

Table 42 shows the summary accuracy statistics. BCWP2 is more accurate overall (MAPE), at the median (Median APE), from 0 to 70% complete, and from 20 to 70% complete. The 20 to 70% completion interval is reported here because this was the interval from the AFCAA study (Keaton, 2014). Overall, the BCWP2 model displayed an accuracy improvement of 7.1% over the reported EAC and 6.5% over BCWP1. Figure 33 shows a visual depiction of the MAPE from the final reported EAC at 10% time intervals. BCWP2 is the most accurate model from contract initiation to approximately 80% complete. BCWP1 experiences an uptick at the 60% mark. A deeper analysis discovered the WGS Block 2 contract was the reason for BCWP1's uptick. BCWP1 uses the CPR PD as its duration estimate. In WGS Block 2, at roughly the 50%

completion point, the CPR PD begins to drastically overestimate the contract's duration. Figure 34 shows the effect of truncating WGS Block 2 at the 50% point. BCWP 1 exhibits a better behaved trajectory once WGS block 2 is truncated. Rather than using a potentially inaccurate CPR PD, the risk could be mitigated by simply using the IMS PD in the BCWP1 model.

Table 42: Accuracy Summary for EAC Forecasting Methods

Metric	Reported EAC	BCWP1	BCWP2	Reported EAC Delta	BCWP1 Delta
MAPE	25.0%	24.4%	17.9%	7.1%	6.5%
Median APE	24.2%	21.2%	17.0%	7.2%	4.2%
MAPE (0 to 70%)	32.9%	28.5%	21.0%	11.9%	7.6%
MAPE (20 to 70%)	28.6%	24.8%	16.3%	12.3%	8.6%

Table 43 shows the EAC accuracy results for individual contracts; BCWP2 is more accurate than the reported EAC in 13/15 contracts and more accurate than BCWP1 in 14/15 contracts. Logically, when the CPR PD estimate is more accurate we would expect the BCWP1 to be more accurate than BCWP2 because BCWP1 uses CPR PD as its duration estimate. An interesting phenomenon occurred in the MUOS-2 and EELV contracts. The CPR PD was the more accurate duration estimate for these two contracts; however, BCWP2 was the more accurate cost estimate compared to the reported EAC and BCWP2. Why did this occur? Time constraints were an obstacle to providing a satisfactory explanation therefore further research is needed to investigate the relationship between duration accuracy and EAC accuracy with the BCWP model.

Table 43: EAC Forecasting Accuracy – Individual Contracts

Contract	Final Duration MAPE		Final EAC MAPE		
	CPR PD	IMS PD /SPI(t)* CPI	Reported EAC	BCWP1	BCWP2
GPS MUE-1	33.0%	**25.0%**	35.6%	29.5%	**19.9%**
GPS MUE-3	32.8%	**28.5%**	37.9%	25.2%	**22.0%**
GPS MUE-4	22.7%	**21.0%**	31.6%	21.2%	**8.7%**
GPS OCX -1	20.4%	**19.9%**	13.9%	13.1%	**12.4%**
GPS OCX- 3	22.7%	**22.0%**	15.8%	16.4%	**15.0%**
WGS B2FO	**29.3%**	36.2%	**2.7%**	25.9%	17.2%
WGS Block 2	24.8%	**20.8%**	17.6%	45.2%	**17.0%**
MUOS-1	**20.3%**	34.4%	**24.2%**	37.1%	28.8%
AEHF	25.7%	**23.1%**	31.6%	20.3%	**16.9%**
SBIRS	24.7%	**24.0%**	39.8%	**31.0%**	31.4%
FAB-T	8.3%	**3.6%**	25.9%	18.0%	**12.2%**
MUOS-2	**8.6%**	9.6%	22.5%	19.9%	**18.5%**
EELV	**5.7%**	9.0%	23.7%	16.8%	**14.4%**
MGUE	23.0%	**15.3%**	16.5%	20.6%	**14.9%**
GPS OCX B	21.0%	**15.1%**	35.4%	24.9%	**18.6%**

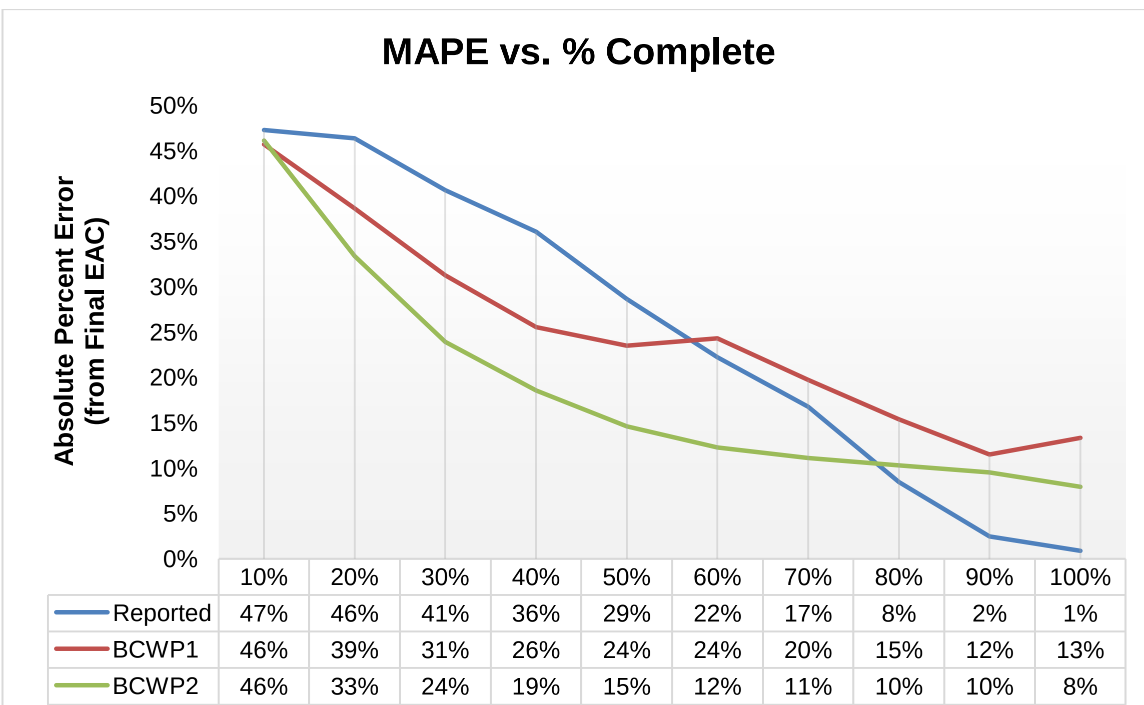


Figure 33: MAPE for EAC Forecasting Methods vs. % Complete

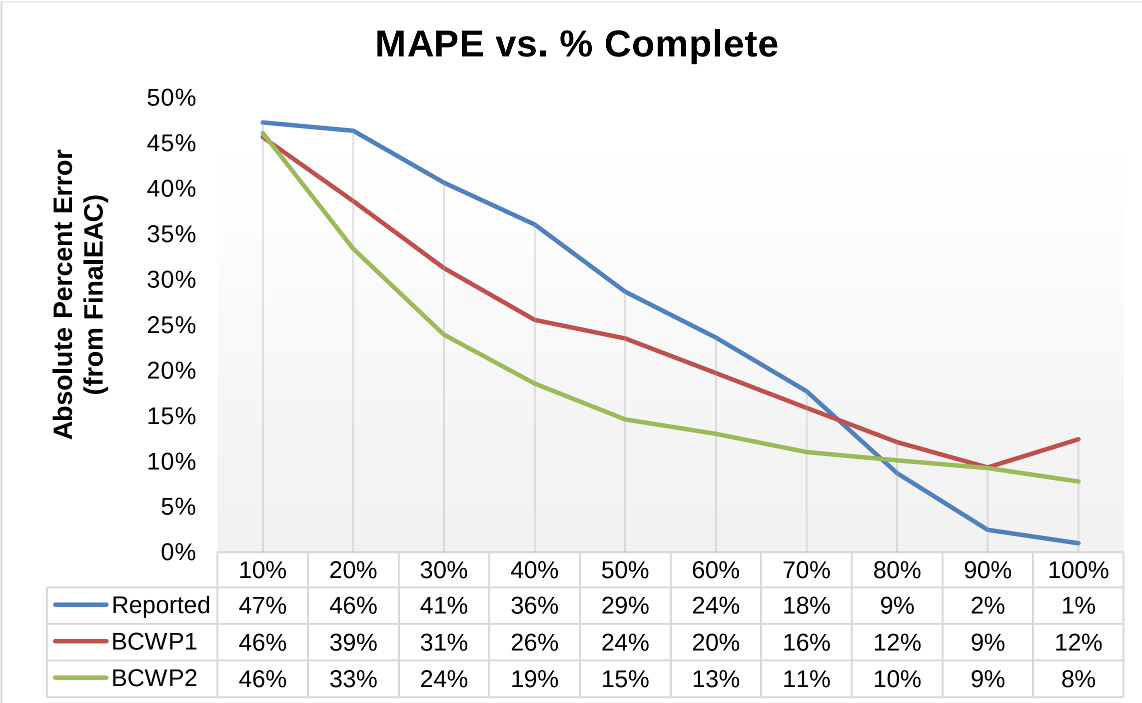


Figure 34: MAPE for EAC Forecasting Methods vs. % Complete [Truncated WGS]

The final analysis is an attempt to provide a more tangible explanation of cost estimate accuracy, this accuracy metric is in dollars rather than MAPE. We converted the mean absolute percent errors (MAPEs) into an average estimating error in dollars. The MAPE for each contract and cost estimating model were multiplied by the final EAC (converted to FY15$). For reference, the total final EAC portfolio cost was $25.7B (FY15$). Figure 35 displays the average cost estimating error for the three models; both the BCWP1 and BCWP2 outperform the EAC. BCWP2 outperforms the EAC by $1.73B and BCWP1 by $0.82B or ($820 million). We caution that these funds are not necessarily savings or potential realizable savings. The BCWP2 model would have provided a more accurate cost estimate to the tune of $820 million (on average) for this portfolio.

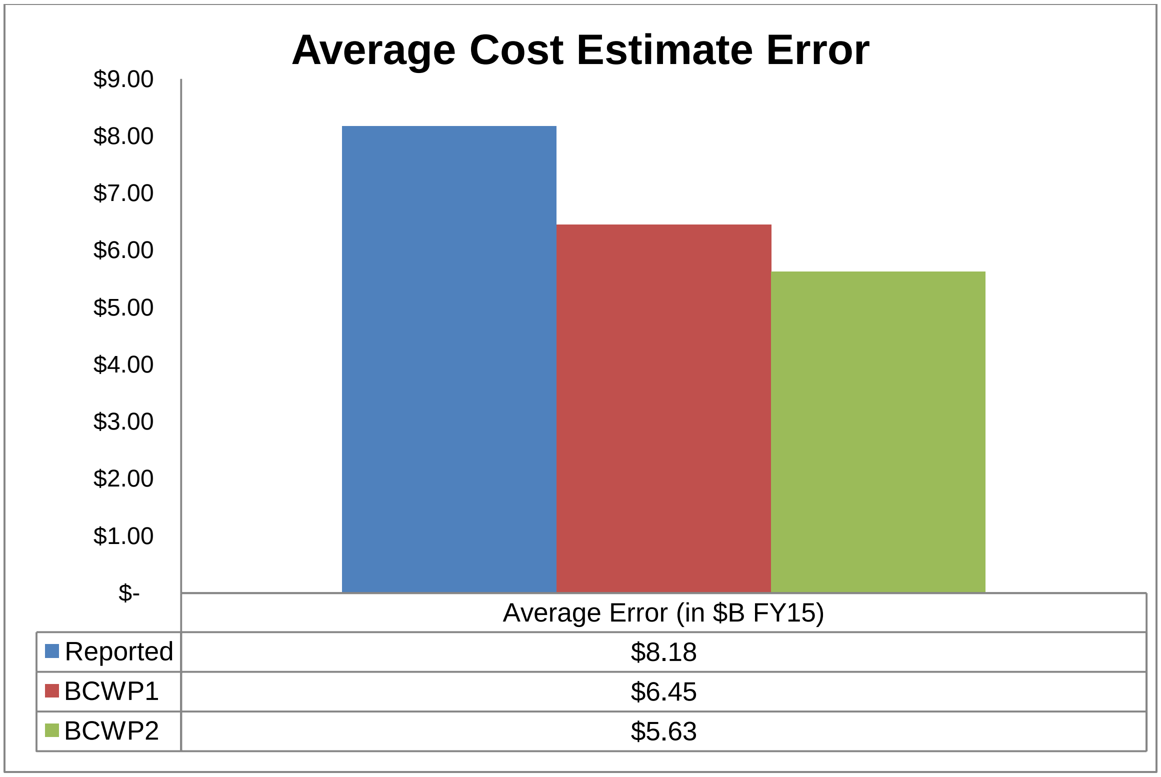


Figure 35: Average Cost Estimate Error (in $B FY15)

Summary

Many of the models reported in this chapter demonstrated improved accuracy over the status quo estimating method, particularly the IDE models. The models were accurate for both OTB and non-OTB contracts. However, short duration contracts without OTBs did not display significantly different results than the status quo. The results were significant for long duration contracts, but less pronounced (alpha = 0.10) than the medium duration contracts (alpha = 0.05). Our regression analysis showed OTBs, schedule growth, and cost growth affected the accuracy of the models. In regards to timeliness, the improvement is most substantial up to the 80% completion point; the accuracy improvement is greater when IDE data is available. For both duration data sets

(IDE and non IDE) the IMS PD is the most accurate model from 80% to 100% completion.

One duration model [IMS PD / (SPI(t)*CPI)] was tested and validated for accuracy in the BCWP burn rate model. The BCWP2 model proved more accurate than the reported EAC and BCWP1 model. The next chapter discusses the policy implications from these results, recommendations, and future research avenues.

V. Conclusions and Recommendations

Investigative Questions Answered

The overall research objective is to evaluate forecasting methods for space contract duration based on the following criteria: accuracy, reliability, and timeliness. In support of the overarching research objective, the following questions were investigated. Our first question was, "What are the appropriate methods to estimate a program's duration?" The methods from the literature include index based, regression, Kalman-Filter, and IMS analysis (to develop IDEs). The new contributions of this research are the addition of the BEI and time series analysis to the index based approach, the Kalman-Filter application to DoD programs, and applying the IMS analysis to space programs.

Our second question was, "How should accuracy be measured and how accurate are the various schedule estimating methods (individual contract, overall, and by various groupings)?" This question represented the bulk of the research. Many accuracy measures were researched, but the MAPE was selected for its applicability across sample sizes and ease of communicating the results. In regards to accuracy, no single model was dominant across all contracts. Of note, the Kalman Filter method did not achieve significant improvements over the status quo and the regression approach was the worst performing model overall. Therefore these methods, as researched here, should be eliminated from consideration. The IDE based models are the most accurate. Combining IDE with the SPI and SPI(t) based performance factors further enhances the accuracy. This analysis shows that the best IDE model is 5.2% more accurate than the status quo (Table 20). If IDE data is not available the best IMS PD model [IMS PD / SPI(t)

(T.S.)*BEI] offered a modest, but significant 2.63% improvement (Table 19). The duration estimating models did not demonstrate significantly different accuracy (compared to the status quo) for short duration contracts. Unfortunately, one limitation of this analysis was the lack of IDE data for the short duration contracts. Medium duration contracts had the largest improvement at 7.80% (Table 25). Of note, each of the medium duration contracts had IDE data. The long duration contracts were significantly different (alpha = 0.10) than the status quo, but the difference was less pronounced than the medium duration contracts. Finally, regression analysis conducted on the model accuracy detected correlation between OTBs, schedule growth, and cost growth. Contracts with one OTB, low schedule growth, and/or low cost growth were correlated with increased accuracy.

Our third question was, “At what point in time (if at all) are the new techniques more accurate than the status quo?” In regards to timeliness, the improvement is most pronounced up to the 80% completion point and the accuracy improvement is greater when IDE data is available. The most pessimistic forecast models were accurate early on (0% to 60%). As the contracts matured (61 to 80%), moderately pessimistic models were more accurate. For both data sets (IDE and non IDE) the IMS PD is the most accurate model from 80% to 100% completion.

Our fourth and last question was, “Are the forecasts accurate for programs with one or more over target baseline (OTB)?” The forecast models offer improved accuracy for programs with OTBs. In fact, the forecasts for OTB programs improve the accuracy (over the status quo) by a larger margin than non OTB programs (3.17% vs. 2.16%). The hypothesis is contracts with OTBs may improve the fidelity of their schedule data compared to non-OTB contracts. Undergoing one OTB seems to be beneficial. However, undergoing multiple OTBs did not improve the duration estimate accuracy.

The genesis of this research was to gauge the accuracy of the status quo method and if possible, improve upon that method. The next step was to determine when (if at all) the accuracy improves over the status quo, and finally, if the models were accurate for OTB contracts. We can definitively conclude that relying on the CPR reported ECD (status quo) is not the best course of action. In fact, simply verifying the dates reported in the IMS is a more accurate method (25.77% compared to 26.14%). Using the IMS PD and EVM indices resulted in a 2.93% accuracy improvement. The potential exists for a larger accuracy improvement (5.2%) when IDE data is available. IMS PD/PF and IDE models are more accurate than the status quo up to the 80% completion point, past this point the accuracy advantage fades. Time series analysis improved accuracy, but not by a significant amount. The Kalman Filter method did not improve accuracy over the status quo. Finally, the regression approach was by far the least accurate model.

A late addition to this research was the validation of the BCWP based cost estimate model. One duration model [IMS PD / (SPI(t)*CPI)] combined with the BCWP burn rate model (BCWP2) outperformed the standard BCWP model (BCWP1) on each accuracy metric. BCWP2 outperforms BCWP1 from 0 to 100% complete. Furthermore, BCWP2 outperforms the reported EAC from 0 to 80% completion.

Recommendations

This research found multiple methods that improve the accuracy of duration estimating for space and development contracts. The improved duration estimates can be used with the BCWP burn rate cost estimate model to further improve the accuracy of cost estimates. Additionally, program managers can take corrective action sooner because the IMS and IDE models exhibit accuracy gains up to the 80% completion point.

Three IDE methods are recommended if IDE data is available: IDE, IDE/SPI, and IDE/SPI(t). One disadvantage associated with developing the IDE models is the process is not as simple as using the IMS PD and performance factors. An additional obstacle is the IDE methodology is relatively new, therefore it will probably not be an accepted best practice for some time. If IDE data does not exist, the IMS PD/ (SPI(t) * BEI) model is recommended because of its simplicity and accuracy. Because they did not offer significant improvement, models with time series based performance factors are not recommended unless the user has access to software comparable to JMP® 11.

Finally, the BCWP2 cost estimate model was validated with fifteen space contracts. This model is recommended because it provided substantial accuracy improvement over both the reported EAC and the BCWP1 model. At a minimum, the BCWP2 model should be used as a cross check for other cost estimating methods.

Recommendations for Future Research

A variety of future research avenues exist. The schedule research was conducted on space and development contracts. Expanding the data set to other commodity and contract types is a logical first step. Another logical step is to test the combination of the AFCAA study's cost model (BCWP1) and additional duration models from this research. Additional research opportunities are derived from fine-tuning the methodology. First, the prediction intervals from the Kalman Filter and time series analysis could be used to develop optimistic and pessimistic forecasts. Restricting the time series analysis to a shorter time frame, for example using 12 months of data at a time, would give more weight to recent performance. Additionally, the OTBs could be incorporated into the time series analysis instead of resetting the analysis after each OTB. In regards to regression, two approaches should be considered: obtaining more data to discover new schedule estimating relationships (SERs) or using current SERs to build a regression model. This regression model could be used to develop an initial duration estimate, then techniques from this research could be used to enhance the duration estimate with EVM data.

Appendix A: Data Adjustments

Table 44: Data Adjustments - AEHF (F04701-02-C-0002)

Report Date	Completion Date	ECD	Adjustment
2/23/2003	1/25/2009		Used reported completion date (1/25/09)
3/30/2003	1/25/2009		Used reported completion date (1/25/09)
8/31/2003	1/25/2009		Used reported completion date (1/25/09)
12/30/2007	5/31/2011		Used reported completion date (5/31/11)
9/25/2011	12/31/2013		Used reported completion date (12/31/13)
4/29/2012	9/30/2013		Used reported completion date (9/30/13)

Table 45: Program: GPS OCX (FA8807-08-C-0001)

Report Date	Completion Date	ECD	Adjustment
12/28/2007	4/30/2009		Used the reported completion date for ECD (4/30/09)
2/1/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)
2/29/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)
3/28/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)
5/2/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)
5/30/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)
6/27/2008	5/30/2009		Used the reported completion date for ECD (5/30/09)

Table 46: Program: GPS OCX (FA8807-08-C-0003)

Report Date	Start Date	ECD	Adjustment
3/28/2010	2/25/2010	3/31/2016	Did not use this month's data. It appears to be from a different contract: different contract start date from the other data points (11/21/07)

Table 47: Months with Missing IDE Data MUOS (N00039-04-C-2009)

Report Date
2/22/2009
3/29/2009
4/26/2009
2/24/2013
3/31/2013
4/28/2013
5/26/2013
6/30/2013
7/28/2013
8/25/2013

Table 48: Data Adjustments - NAVSTAR GPS (FA8807-06-C-0001)

Report Date	ECD	Adjustment
7/28/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
9/1/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
9/29/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
10/27/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
12/1/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
12/29/2006		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
2/2/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
3/2/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
3/30/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
4/27/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
6/1/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
6/29/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
7/27/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
8/31/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
9/28/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
11/2/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
11/30/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
12/28/2007		Used IMS reported completion date (11/2/09) from first IMS 2/20/08
2/1/2008		Used IMS reported completion date (11/2/09) from first IMS 2/20/08

Table 49: Months with Missing IDE Data - NAVSTAR GPS (FA8807-06-C-0001)

Report Date
7/28/2006
9/1/2006
9/29/2006
10/27/2006
12/1/2006
12/29/2006
2/2/2007
3/2/2007
3/30/2007
4/27/2007
6/1/2007
6/29/2007
7/27/2007
8/31/2007
9/28/2007
11/2/2007
11/30/2007
12/28/2007
2/1/2008
2/29/2008
5/30/2008
6/27/2008
8/1/2008
12/3/2010
12/31/2010
1/28/2011
2/25/2011
4/1/2011
4/29/2011
6/3/2011
7/1/2011
7/29/2011
9/2/2011
9/30/2011
3/30/2012
2/1/2013
3/1/2013
3/29/2013

Table 50: Data Adjustments - NAVSTAR GPS (FA8807-06-C-0001)

Report Date	BAC	BCWS	BCWP	ACWP	Notes
5/30/2008	71,098,080,040	43,439,985,020	40,962,631,470	45,463,954,240	Inconsistent values. Verified amount should be millions (July 2013 Format 3).
6/27/2008	71,098,077,750	45,894,059,080	43,977,857,780	48,645,986,660	Corrected values (divided by 1000)
8/29/2008	72,389,214,760	50,872,327,110	50,602,492,380	54,705,923,800	Corrected values (divided by 1000)
11/7/2008	72,434,559,250	57,014,389,560	56,527,728,370	61,614,545,870	Corrected values (divided by 1000)
1/2/2009	72,736,320,630	60,263,855,720	59,003,046,410	64,426,448,100	Corrected values (divided by 1000)
1/30/2009	74,290,042,550	61,257,442,270	60,301,225,820	66,099,221,580	Corrected values (divided by 1000)
2/27/2009	74,253,744,010	62,243,481,210	61,407,136,860	67,728,083,900	Corrected values (divided by 1000)
4/3/2009	74,746,177,020	63,523,323,150	62,653,351,860	69,959,445,080	Corrected values (divided by 1000)
5/29/2009	75,916,157,950	66,466,698,820	64,821,227,730	72,251,147,020	Corrected values (divided by 1000)
7/31/2009	76,049,327,110	68,531,432,140	67,633,591,370	75,258,526,470	Corrected values (divided by 1000)
8/28/2009	76,077,136,590	70,424,503,960	69,449,895,160	77,092,417,090	Corrected values (divided by 1000)
1/1/2010	76,017,437,990	73,275,299,700	72,369,324,180	81,844,861,410	Corrected values (divided by 1000)
1/29/2010	76,023,959,900	73,592,144,180	72,759,835,470	83,188,966,060	Corrected values (divided by 1000)
2/26/2010	75,667,517,120	73,678,546,600	72,992,968,390	84,534,934,370	Corrected values (divided by 1000)
4/30/2010	75,667,511,550	74,123,735,390	73,359,370,260	87,256,865,430	Corrected values (divided by 1000)
7/2/2010	26,397,443	26,397,443	26,397,568	30,025,268	Did not use this month's data. Data appears to be from different contract.
7/30/2010	75,721,045,370	74,308,069,770	73,973,581,820	90,033,758,670	Corrected values (divided by 1000)
8/27/2010	75,721,045,370	74,358,623,460	74,216,251,400	90,603,357,170	Corrected values (divided by 1000)
12/3/2010	75,721,047,790	74,384,289,790	74,382,668,930	91,484,071,110	Corrected values (divided by 1000)
12/31/2010	75,721,048,970	74,384,289,790	74,382,668,930	91,573,441,170	Corrected values (divided by 1000)
1/28/2011	75,721,048,970	74,384,290,970	74,384,413,000	91,783,913,030	Corrected values (divided by 1000)
2/25/2011	75,721,048,970	74,384,290,970	74,384,413,000	91,953,844,170	Corrected values (divided by 1000)
4/1/2011	75,721,049,010	74,384,293,380	74,384,415,410	92,093,863,310	Corrected values (divided by 1000)
4/29/2011	75,721,049,010	74,384,291,010	74,384,414,220	92,337,336,560	Corrected values (divided by 1000)
6/3/2011	75,721,049,010	74,384,291,010	74,384,414,220	92,398,331,150	Corrected values (divided by 1000)
7/29/2011	75,721,050,220	74,384,291,010	74,384,414,220	92,429,535,010	Corrected values (divided by 1000)
9/30/2011	111,254,427,970	80,334,888,200	79,646,618,390	97,153,544,210	Corrected values (divided by 1000)
3/30/2012	121,909,362,420	103,471,362,310	103,728,019,060	107,298,090,770	Corrected values (divided by 1000)
4/27/2012	122,029,968,270	104,860,494,760	104,932,474,730	108,230,414,990	Corrected values (divided by 1000)
6/1/2012	122,102,957,260	106,700,407,470	106,483,744,560	109,668,903,710	Corrected values (divided by 1000)
6/29/2012	121,953,566,350	107,996,524,900	108,043,514,270	111,240,130,270	Corrected values (divided by 1000)
8/3/2012	122,300,124,770	109,718,335,580	109,193,583,820	112,705,700,500	Corrected values (divided by 1000)
8/31/2012	122,259,444,490	110,894,265,980	110,686,336,810	114,634,671,950	Corrected values (divided by 1000)
9/28/2012	121,917,682,130	112,802,176,580	112,884,996,410	116,898,310,100	Corrected values (divided by 1000)
11/30/2012	122,093,327,350	114,728,468,640	114,958,572,800	118,866,848,350	Corrected values (divided by 1000)
2/1/2013	122,063,787,150	117,242,810,500	116,642,266,010	120,415,610,630	Corrected values (divided by 1000)
3/1/2013	123,543,730,230	117,907,830,010	117,514,885,230	121,381,259,310	Corrected values (divided by 1000)
3/29/2013	123,387,178,960	118,400,041,350	118,104,891,740	121,999,091,410	Corrected values (divided by 1000)
5/3/2013	123,555,159,450	119,429,196,680	119,349,668,270	123,132,402,880	Corrected values (divided by 1000)
5/31/2013	123,578,144,670	120,241,157,130	119,934,629,250	123,684,311,600	Corrected values (divided by 1000)
6/28/2013	23,530,727,520	10,179,436,740	10,118,731,740	10,051,101,890	Did not use this month's data. Data appears to be from different contract: different start date (9/28/12 vs. 5/26/06)
8/2/2013	123,625,515,410	121,876,840,320	121,406,668,960	125,104,405,170	Corrected values (divided by 1000)

Table 51: Data Adjustments - NAVSTAR GPS (FA8807-06-C-0003)

Report Date	Completion Date	ECD	Adjustment
11/24/2006	10/30/2007		Used the reported Completion date of 10/30/07
12/29/2006	10/30/2007		Used the reported Completion date of 10/30/07
1/26/2007	10/30/2007		Used the reported Completion date of 10/30/07
2/23/2007	10/30/2007		Used the reported Completion date of 10/30/07
3/30/2007	10/30/2007		Used the reported Completion date of 10/30/07
4/27/2007	10/30/2007		Used the reported Completion date of 10/30/07
5/25/2007	10/30/2007		Used the reported Completion date of 10/30/07
6/29/2007	10/30/2007		Used the reported Completion date of 10/30/07

Table 52: Months with Missing IDE Data - NAVSTAR GPS (FA8807-06-C-0003)

Report Date
11/24/2006
12/29/2006
1/26/2007
2/23/2007
3/30/2007
4/27/2007
5/25/2007
6/29/2007
7/27/2007
8/24/2007
9/28/2007
10/26/2007
11/23/2007
12/28/2007
1/25/2008
2/22/2008
3/28/2008
4/25/2008
12/31/2008
2/25/2011
4/1/2011
4/29/2011
5/27/2011
7/1/2011
7/29/2011
8/26/2011
9/30/2011
10/28/2011
11/25/2011
12/30/2011
1/27/2012
2/24/2012
3/30/2012
4/27/2012
5/25/2012
6/29/2012
7/27/2012
8/24/2012
9/28/2012

Table 53: Data Adjustments - NAVSTAR GPS (FA8807-06-C-0004)

Report Date	Start Date	ECD	Adjustment
11/18/07	6/26/06	12/31/07	Did not use data from this month. It appears to be from a different contract: different start date (6/26/06 vs. 6/02/06).
12/31/07	6/02/06		Used next month's ECD (1/12/11).

Table 54: Months with Missing IDE Data - NAVSTAR GPS (FA8807-06-C-0004)

Report Date
12/31/2007
1/27/2008
2/24/2008
5/25/2008
8/24/2008
6/28/2009
7/25/2010
2/26/2012
10/26/2012

Table 55: Months with Missing IDE Data – WGS Blk 2 (FA8808-06-C-0001)

Report Date
11/30/2006
12/21/2006
1/25/2007
2/22/2007
3/29/2007
4/26/2007
5/31/2007
6/28/2007
7/26/2007
8/30/2007
9/27/2007
10/25/2007
11/29/2007
12/20/2007
1/31/2008
2/28/2008
3/27/2008
4/24/2008
4/26/2012
5/31/2012
6/28/2012
7/26/2012
8/30/2012
9/27/2012
10/25/2012
11/29/2012
12/20/2012
1/31/2013
2/28/2013
3/28/2013
4/25/2013
5/30/2013
6/27/2013
7/25/2013
8/29/2013
9/26/2013
10/31/2013
11/28/2013
12/19/2013
1/30/2014

Table 56: Months with Missing IDE Data - WGS B2FO (FA8808-10-C-0001)

Report Date
10/28/2010
11/25/2010
12/23/2010
11/28/2013
12/19/2013
1/30/2014
2/27/2014
3/27/2014
4/24/2014

Table 57: Additional Data - SBIRS (F04701-95-C-0017)

Notes
Additional data (from 12/1/96 until 7/26/2004) was provided by the author of the AFCAA research (Keaton, 2014).

Table 58: Data Adjustment – SBIRS (F04701-95-C-0017)

Report Date	Original BAC	Prior BAC	Adjusted BAC	Next BAC	Adjustment
8/29/04	3,311,589,000	5,259,883,000	5,274,890,122	5,317,410,300	Adjusted BAC with linear interpolation for regression forecast.
1/29/06	4,206,867,200	5,675,887,300	5,920,741,440	6,173,757,384	Adjusted BAC with linear interpolation for regression forecast.
12/30/07	5,414,927,378	6,555,123,944	6,725,182,546	6,906,578,389	Adjusted BAC with linear interpolation for regression forecast.

Table 59: Months with Missing IDE Data – AEHF (F04701-02-C-0002)

Report Date			
1/1/1997	1/1/2000	1/1/2003	1/29/2006
2/1/1997	2/1/2000	2/1/2003	2/26/2006
3/1/1997	3/1/2000	3/1/2003	3/26/2006
4/1/1997	4/1/2000	4/1/2003	4/30/2006
5/1/1997	5/1/2000	5/1/2003	5/28/2006
6/1/1997	6/1/2000	6/1/2003	6/25/2006
7/1/1997	7/1/2000	7/1/2003	7/30/2006
8/1/1997	8/1/2000	8/1/2003	8/27/2006
9/1/1997	9/1/2000	9/1/2003	9/24/2006
10/1/1997	10/1/2000	10/1/2003	10/29/2006
11/1/1997	11/1/2000	11/1/2003	11/26/2006
12/1/1997	12/1/2000	12/1/2003	12/31/2006
1/1/1998	1/1/2001	1/1/2004	1/28/2007
2/1/1998	2/1/2001	2/1/2004	2/25/2007
3/1/1998	3/1/2001	3/1/2004	3/25/2007
4/1/1998	4/1/2001	4/1/2004	4/29/2007
5/1/1998	5/1/2001	5/1/2004	5/27/2007
6/1/1998	6/1/2001	6/1/2004	6/24/2007
7/1/1998	7/1/2001	7/1/2004	7/29/2007
8/1/1998	8/1/2001	8/29/2004	8/26/2007
9/1/1998	9/1/2001	9/26/2004	9/30/2007
10/1/1998	10/1/2001	10/31/2004	10/28/2007
11/1/1998	11/1/2001	11/28/2004	11/25/2007
12/1/1998	12/1/2001	12/26/2004	12/30/2007
1/1/1999	1/1/2002	1/30/2005	1/27/2008
2/1/1999	2/1/2002	2/27/2005	2/24/2008
3/1/1999	3/1/2002	3/27/2005	3/30/2008
4/1/1999	4/1/2002	4/24/2005	4/27/2008
5/1/1999	5/1/2002	5/29/2005	5/25/2008
6/1/1999	6/1/2002	6/26/2005	6/29/2008
7/1/1999	7/1/2002	7/31/2005	
8/1/1999	8/1/2002	8/28/2005	
9/1/1999	9/1/2002	9/25/2005	
10/1/1999	10/1/2002	10/30/2005	
11/1/1999	11/1/2002	11/27/2005	
12/1/1999	12/1/2002	12/25/2005	

Table 60: Data Adjustment – WGS Block 2 (FA8808-06-C-0001)

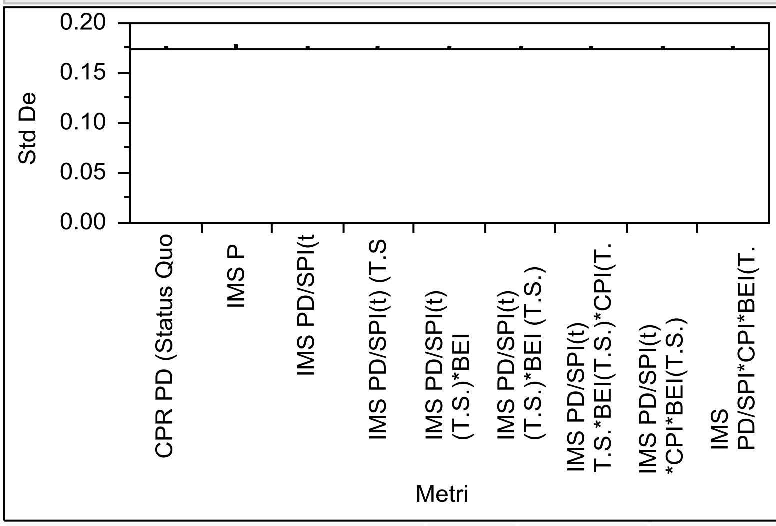
Report Date
11/30/2006
12/21/2006
1/25/2007
2/22/2007
3/29/2007
4/26/2007
5/31/2007
6/28/2007
7/26/2007
8/30/2007
9/27/2007
10/25/2007
11/29/2007
12/20/2007
1/31/2008
2/28/2008
3/27/2008
4/24/2008
4/26/2012
5/31/2012
6/28/2012
7/26/2012
8/30/2012
9/27/2012
10/25/2012
11/29/2012
12/20/2012
1/31/2013
2/28/2013
3/28/2013
4/25/2013
5/30/2013
6/27/2013
7/25/2013
8/29/2013
9/26/2013
10/31/2013
11/28/2013
12/19/2013
1/30/2014

Table 61: Months with Missing IDE Data - WGS B2FO (FA8808-10-C-0001)

Report Date
10/28/2010
11/25/2010
12/23/2010
11/28/2013
12/19/2013
1/30/2014
2/27/2014
3/27/2014
4/24/2014

Appendix B: Levene Tests for Tukey-Kramer HSD

Tests that the Variances are Equal

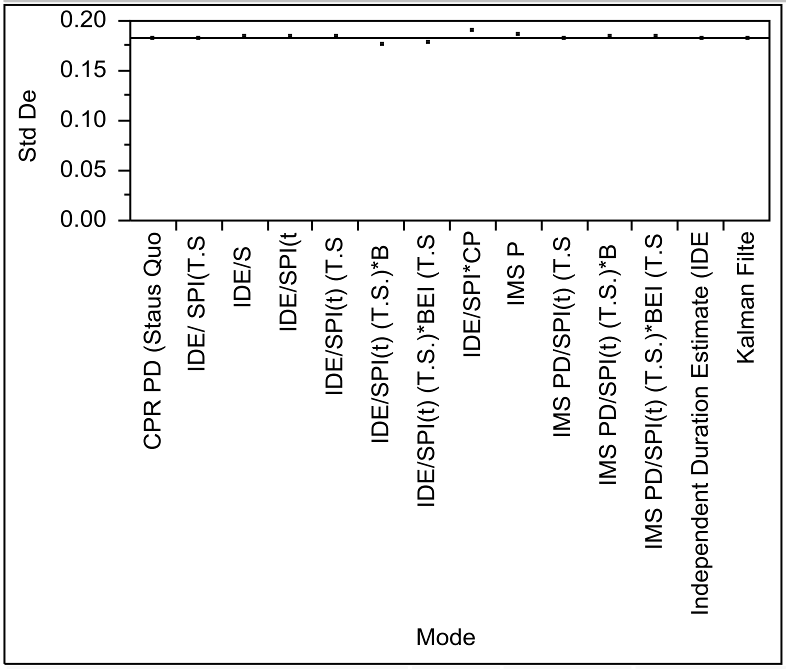


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (Status Quo)	806	0.1730312	0.1456448	0.1446919
IMS PD	806	0.1755445	0.1460916	0.1455804
IMS PD/SPI(t)	806	0.1734472	0.1446154	0.1445243
IMS PD/SPI(t) (T.S.)	806	0.1730525	0.1445527	0.1445527
IMS PD/SPI(t) (T.S.)*BEI	806	0.1730828	0.1462842	0.1461087
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	806	0.1745397	0.1480904	0.1480036
IMS PD/SPI(t) T.S.*BEI(T.S.)*CPI(T.S	806	0.1748482	0.1438965	0.1423427
IMS PD/SPI(t)*CPI*BEI(T.S.)	806	0.1742713	0.1424038	0.1410266
IMS PD/SPI*CPI*BEI(T.S.)	806	0.1740052	0.1422498	0.1409024

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.0435	8	7245	1.0000
Brown-Forsyth	0.4652	8	7245	0.8813
Levene	0.3106	8	7245	0.9624
Bartlett	0.0428	8	.	1.0000

Figure 36: Levene Test (All Contracts)

Tests that the Variances are Equal

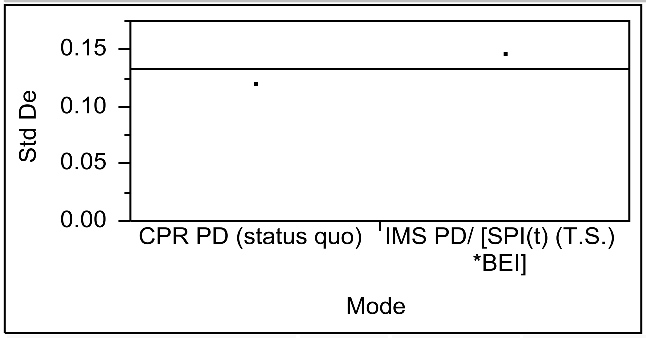


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (Staus Quo)	617	0.1817221	0.1533071	0.1528862
IDE/ SPI(T.S.)	617	0.1824486	0.1509054	0.1482535
IDE/SPI	617	0.1835454	0.1534774	0.1502421
IDE/SPI(t)	617	0.1846534	0.1551278	0.1536426
IDE/SPI(t) (T.S.)	617	0.1837609	0.1535651	0.1527455
IDE/SPI(t) (T.S.)*BEI	617	0.1758765	0.1425820	0.1387472
IDE/SPI(t) (T.S.)*BEI (T.S.)	617	0.1788637	0.1467136	0.1424496
IDE/SPI*CPI	617	0.1899886	0.1585321	0.1571365
IMS PD	617	0.1851526	0.1541202	0.1540206
IMS PD/SPI(t) (T.S.)	617	0.1828119	0.1525956	0.1523511
IMS PD/SPI(t) (T.S.)*BEI	617	0.1830886	0.1552107	0.1544647
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	617	0.1848241	0.1575064	0.1568921
Independent Duration Estimate (IDE	617	0.1818675	0.1503748	0.1485771
Kalman Filter	617	0.1815860	0.1484729	0.1470337

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.3522	13	8624	0.9833
Brown-Forsyth	1.4554	13	8624	0.1259
Levene	1.0975	13	8624	0.3554
Bartlett	0.3777	13	.	0.9771

Figure 37: Levene Test (IDE Contracts)

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (status quo)	175	0.1196852	0.1056706	0.1055749
IMS PD/ [SPI(t) (T.S.)*BE	175	0.1454320	0.1209870	0.1209034

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	11.5163	1	348	0.0008 *
Brown-Forsyth	4.1645	1	348	0.0420 *
Levene	4.3111	1	348	0.0386 *
Bartlett	6.5456	1	.	0.0105 *
F Test 2-sided	1.4765	174	174	0.0105 *

Welch's Test

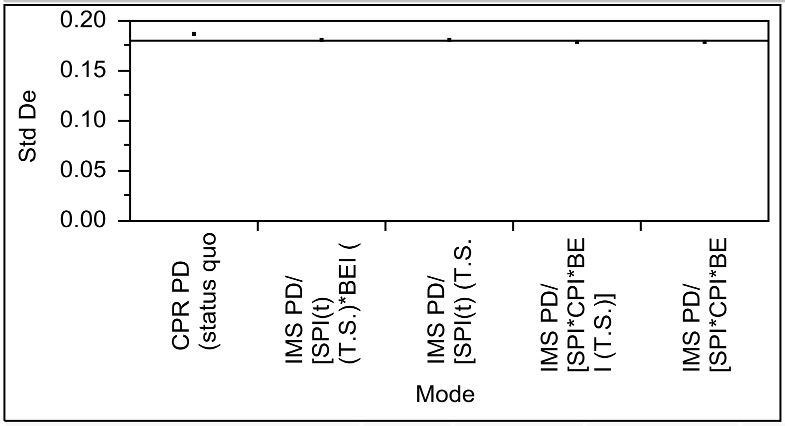
Welch Anova testing Means Equal, allowing Std Devs Not Equ

F Ratio	DFNum	DFDen	Prob > F
2.3019	1	335.58	0.1302

t Test
1.5172

Figure 38: Levene Test (Non-OTB Contracts)

Tests that the Variances are Equal

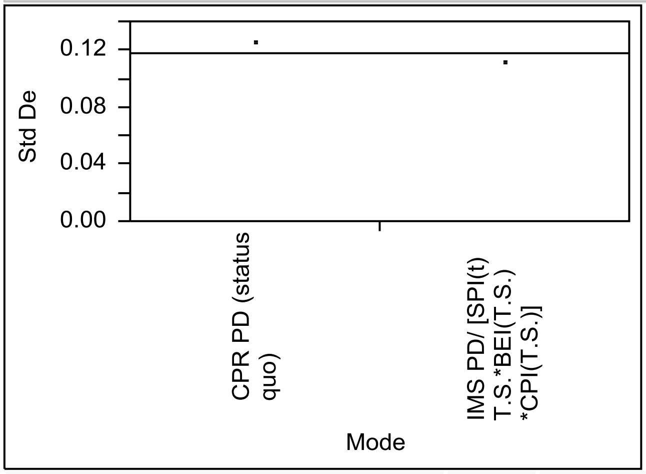


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (status quo)	631	0.1851624	0.1565559	0.1543387
IMS PD/ [SPI(t) (T.S.)*BEI (T.S.	631	0.1804367	0.1535036	0.1526426
IMS PD/ [SPI(t) (T.S.)*BEI]	631	0.1801032	0.1532729	0.1524323
IMS PD/ [SPI*CPI*BEI (T.S.)]	631	0.1777332	0.1454737	0.1423555
IMS PD/ [SPI*CPI*BEI]	631	0.1775436	0.1449411	0.1419983

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.3651	4	3150	0.8336
Brown-Forsyth	1.9989	4	3150	0.0920
Levene	1.7775	4	3150	0.1305
Bartlett	0.3652	4	.	0.8335

Figure 39: Levene Test (OTB Contracts)

Tests that the Variances are Equal

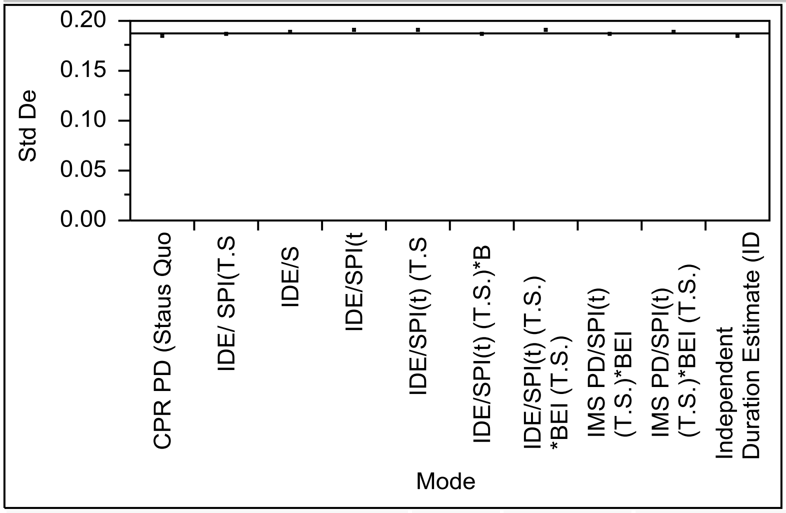


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (status quo)	45	0.1243291	0.1070653	0.0939889
IMS PD/ [SPI(t) T.S.*BEI(T.S.)*CPI(T.S.	45	0.1105880	0.0952114	0.0896911

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	1.1538	1	88	0.2857
Brown-Forsyth	0.0572	1	88	0.8115
Levene	0.9448	1	88	0.3337
Bartlett	0.5954	1	.	0.4403
F Test 2-sided	1.2639	44	44	0.4403

Figure 40: Levene Test - Short Duration (GPS OCX)

Tests that the Variances are Equal

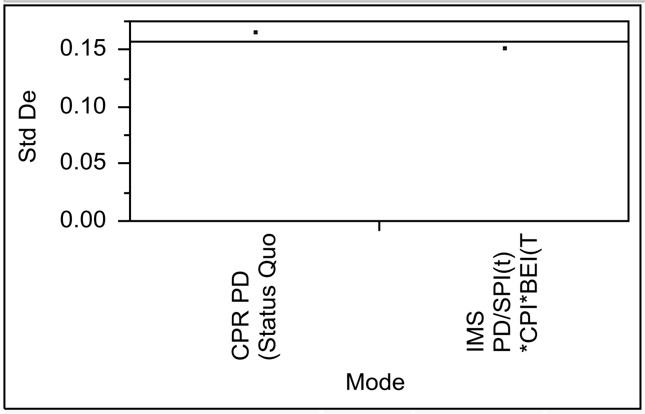


Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (Staus Quo)	405	0.1843967	0.1532158	0.1512249
IDE/ SPI(T.S.)	405	0.1869670	0.1476684	0.1397057
IDE/SPI	405	0.1888509	0.1518587	0.1428074
IDE/SPI(t)	405	0.1903794	0.1547366	0.1476904
IDE/SPI(t) (T.S.)	405	0.1892029	0.1531040	0.1481079
IDE/SPI(t) (T.S.)*BEI	405	0.1858878	0.1487672	0.1441393
IDE/SPI(t) (T.S.)*BEI (T.S.)	405	0.1899459	0.1538366	0.1488180
IMS PD/SPI(t) (T.S.)*BEI	405	0.1859238	0.1507246	0.1504936
IMS PD/SPI(t) (T.S.)*BEI (T.S.)	405	0.1879186	0.1531470	0.1531494
Independent Duration Estimate (IDE	405	0.1849328	0.1461983	0.1398417

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.0694	9	4040	0.9999
Brown-Forsyth	0.5402	9	4040	0.8461
Levene	0.2767	9	4040	0.9811
Bartlett	0.1040	9	.	0.9996

Figure 41: Levene Test - Medium Duration (NAVSTAR GPS, MUOS, & WGS)

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
CPR PD (Status Quo)	356	0.1639028	0.1426005	0.1413340
IMS PD/SPI(t)*CPI*BEI(T.S.	356	0.1501078	0.1275090	0.1234601

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	6.0815	1	710	0.0139 *
Brown-Forsyth	7.2917	1	710	0.0071 *
Levene	6.3842	1	710	0.0117 *
Bartlett	2.7367	1	.	0.0981
F Test 2-sided	1.1922	355	355	0.0981

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equ

F Ratio	DFNum	DFDen	Prob > F
3.2602	1	704.58	0.0714

t Test
1.8056

Figure 42: Levene Test - Long Duration (AEHF & SBIRS)

Appendix C: Duration Accuracy Results (Individual Contracts)

Table 62: NAVSTAR GPS (FA8807-06-C-0001) Accuracy Results

Percent Complete Interval	Forecasting Model: CPR PD (status quo)	IMS PD	IDE	IMS PD / SPI(t) * CPI * BEI	IDE / SPI(t) (T.S.) * BEI	Regress-ion	Kalman Filter
0 to 10	52.72%	52.72%	52.72%	37.76%	49.55%	74.58%	52.72%
11 to 20	52.72%	52.72%	52.72%	42.05%	47.91%	80.11%	52.72%
21 to 30	51.75%	51.75%	51.75%	43.07%	42.10%	63.11%	48.86%
31 to 40	50.26%	50.45%	43.34%	42.26%	40.10%	52.74%	52.42%
41 to 50	47.04%	46.95%	29.00%	36.40%	23.83%	52.29%	46.07%
51 to 60	40.82%	41.84%	17.38%	21.41%	7.72%	53.17%	44.53%
61 to 70	19.57%	19.57%	14.61%	7.03%	6.86%	50.60%	35.93%
71 to 80	11.16%	11.16%	11.16%	5.03%	10.06%	40.89%	27.36%
81 to 90	0.00%	0.00%	8.32%	6.78%	5.07%	15.14%	0.71%
91 to 100	0.00%	0.00%	4.33%	5.56%	6.08%	15.79%	1.20%
MAPE	33.05%	33.16%	29.26%	25.14%	**24.45%**	50.57%	36.44%

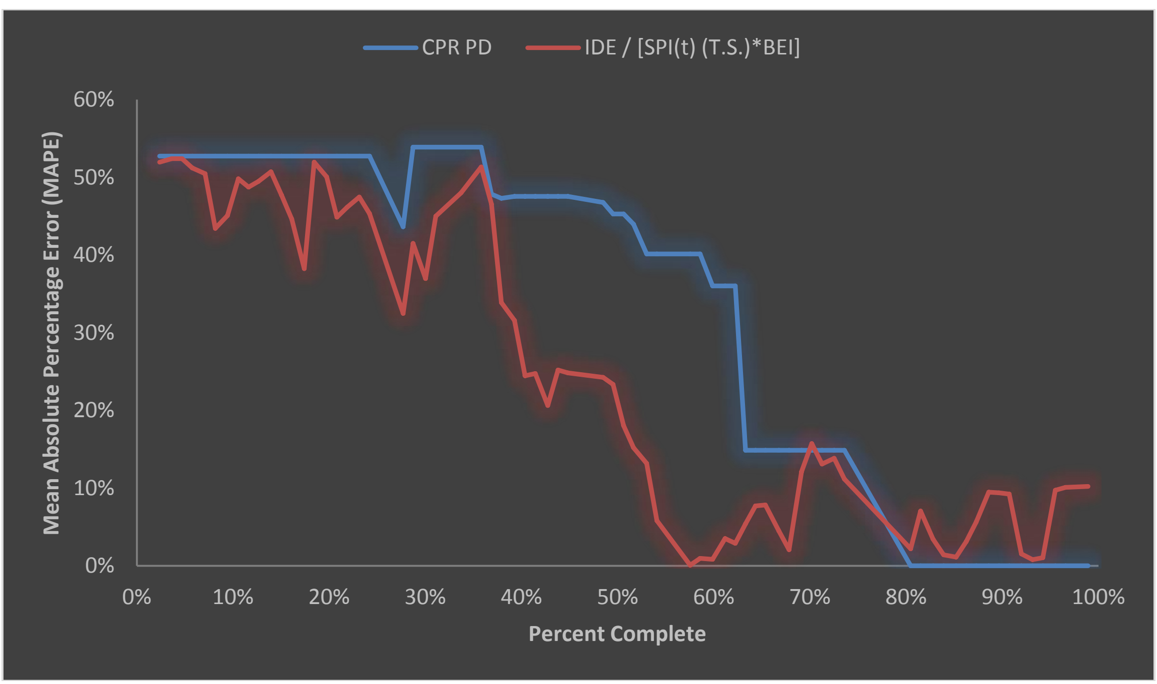


Figure 43: NAVSTAR GPS (FA8807-06-C-0001) Accuracy over Time

Table 63: NAVSTAR GPS (FA8807-06-C-0003) Accuracy Results

Percent Complete Interval	Forecasting Model						
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) T.S. * BEI(T.S.) * CPI(T.S.)]	Regress-ion	Kalman Filter	IDE	IDE / [SPI(t) (T.S.) * BEI (T.S.)]
0 to 10	80.56%	80.56%	80.08%	72.81%	80.56%	80.54%	80.19%
11 to 20	79.66%	79.66%	76.61%	76.22%	79.66%	79.64%	77.60%
21 to 30	41.65%	41.65%	32.70%	71.81%	40.12%	36.37%	29.45%
31 to 40	36.64%	36.31%	25.73%	63.55%	34.22%	28.27%	28.13%
41 to 50	36.72%	36.72%	30.53%	69.15%	34.86%	29.06%	23.17%
51 to 60	36.72%	36.47%	13.71%	69.54%	32.04%	17.10%	7.92%
61 to 70	29.73%	29.73%	11.12%	65.43%	26.16%	20.59%	18.92%
71 to 80	2.42%	4.37%	1.05%	14.11%	2.97%	3.94%	4.02%
81 to 90	0.77%	1.58%	1.32%	22.46%	1.73%	4.83%	6.57%
91 to 100	0.77%	0.19%	1.32%	22.46%	1.73%	4.83%	6.57%
MAPE	32.89%	32.69%	26.14%	56.74%	31.75%	27.91%	**25.67%**

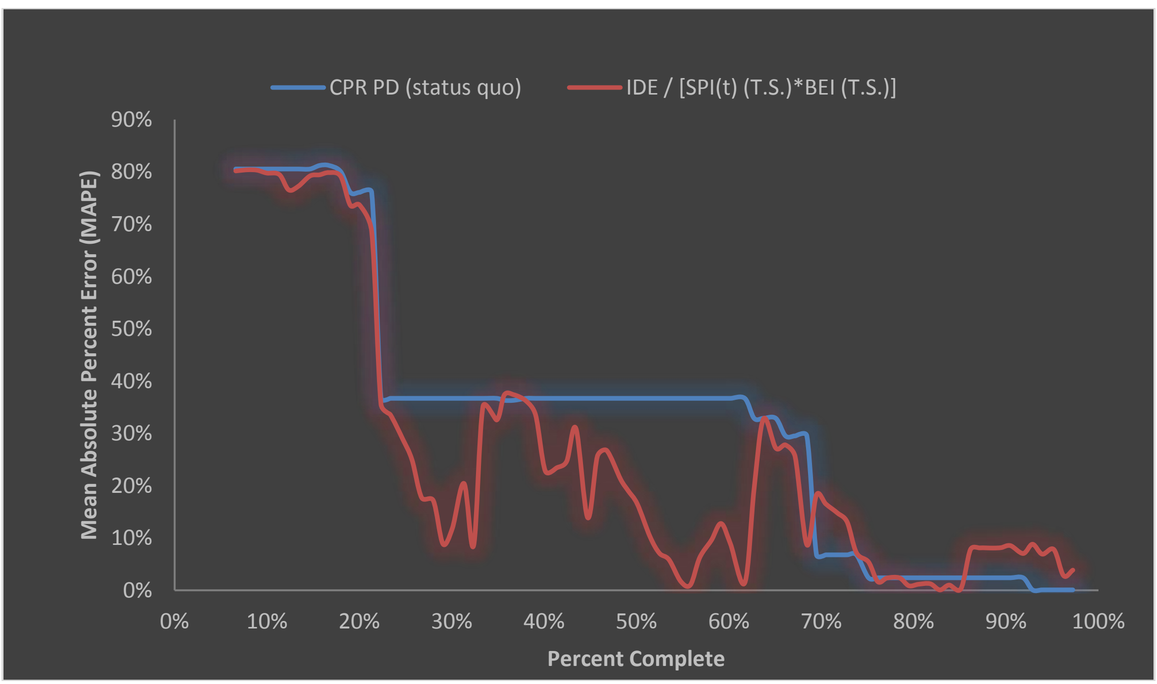


Figure 44: NAVSTAR GPS (FA8807-06-C-0003) Accuracy over Time

Table 64: NAVSTAR GPS (FA8807-06-C-0004) Accuracy Results

Percent Complete Interval	Forecasting Model						
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) (T.S.)*BEI]	Regress-ion	Kalman Filter	IDE	IDE/SPI
0 to 10							
11 to 20							
21 to 30	36.35%	36.35%	35.66%	62.01%	36.68%	27.17%	26.54%
31 to 40	36.29%	35.98%	31.05%	39.65%	35.99%	6.32%	5.43%
41 to 50	36.08%	34.05%	25.71%	49.00%	35.15%	21.04%	18.38%
51 to 60	36.08%	33.08%	9.19%	47.30%	28.47%	17.53%	14.77%
61 to 70	19.62%	29.84%	5.11%	42.25%	16.15%	8.92%	7.49%
71 to 80	12.36%	21.36%	6.40%	28.29%	12.14%	3.23%	3.09%
81 to 90	6.35%	6.16%	4.51%	42.78%	4.35%	5.82%	5.03%
91 to 100	3.18%	2.78%	0.92%	18.64%	1.82%	3.58%	2.22%
MAPE	23.76%	25.59%	14.92%	41.47%	21.75%	11.66%	**10.33%**

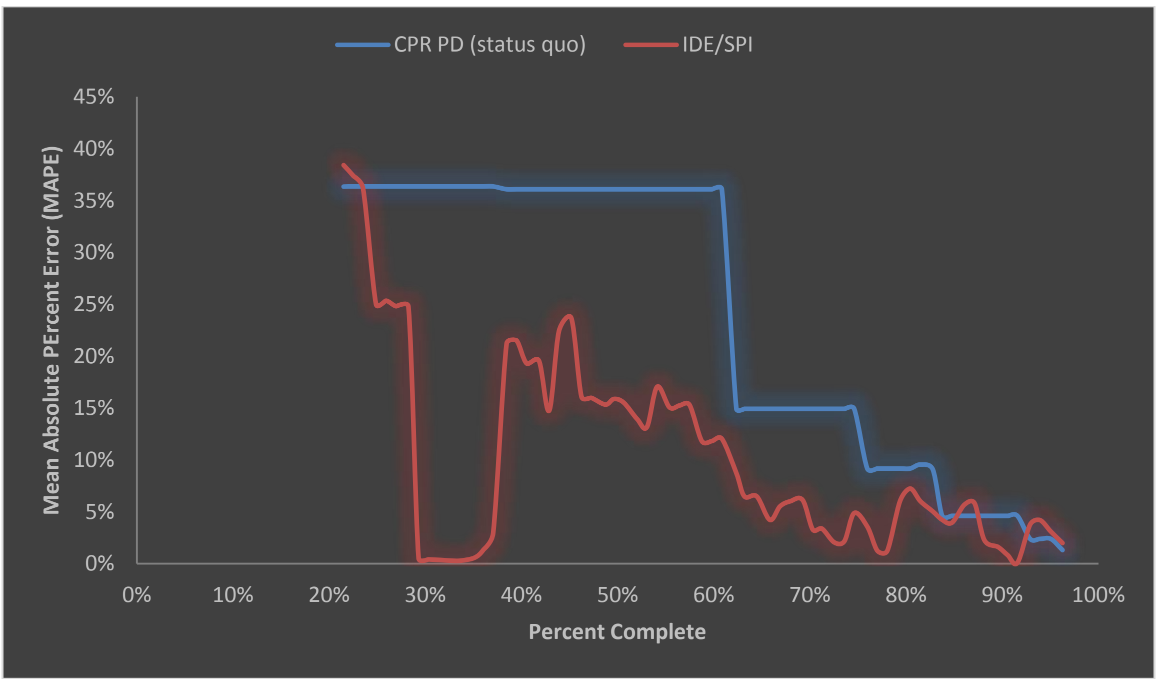


Figure 45: NAVSTAR GPS (FA8807-06-C-0004) Accuracy over Time

Table 65: GPS OCX (FA8807-08-C-0001) Accuracy Results

Percent Complete Interval	Forecasting Model				
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) T.S.* BEI(T.S.)* CPI(T.S.)]	Regression	Kalman Filter
0 to 10	29.74%	30.39%	31.37%	30.87%	28.95%
11 to 20	27.76%	28.95%	25.35%	31.50%	28.94%
21 to 30	27.76%	28.16%	20.92%	32.35%	28.46%
31 to 40					
41 to 50	27.76%	25.79%	19.45%	32.83%	29.03%
51 to 60	27.76%	28.16%	25.57%	29.34%	29.24%
61 to 70	27.76%	28.16%	26.88%	20.85%	30.33%
71 to 80	17.24%	16.18%	15.16%	18.31%	17.18%
81 to 90	5.99%	5.99%	6.29%	16.09%	12.63%
91 to 100	0.00%	0.00%	0.51%	11.41%	0.35%
MAPE	20.41%	20.49%	**18.37%**	24.08%	21.73%

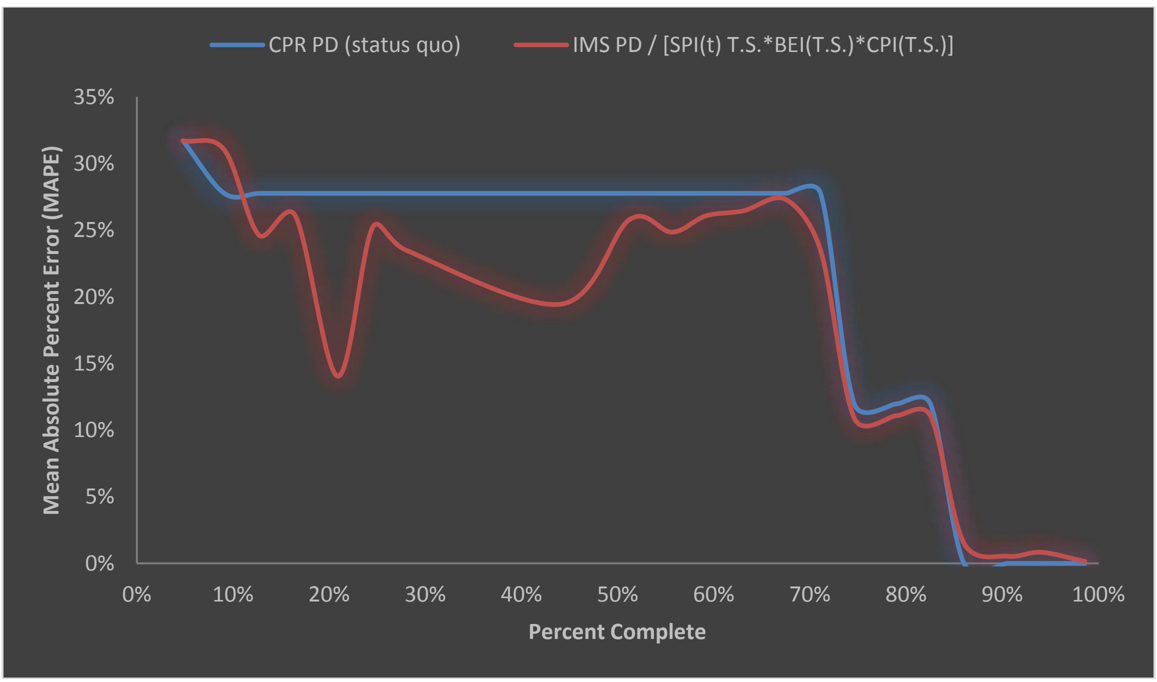


Figure 46: GPS OCX (FA8807-08-C-0001) Accuracy over Time

Table 66: GPS OCX (FA8807-08-C-0003) Accuracy Results

Percent Complete Interval	Forecasting Model				
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) *CPI *BEI]	Regression	Kalman Filter
0 to 10	36.12%	36.47%	33.94%	6.47%	36.47%
11 to 20	35.76%	36.47%	26.63%	26.47%	36.46%
21 to 30	35.76%	35.41%	20.84%	22.33%	35.42%
31 to 40	35.76%	33.53%	29.64%	35.90%	34.58%
41 to 50	35.76%	34.51%	34.05%	40.65%	35.35%
51 to 60	28.24%	31.71%	31.72%	33.43%	31.48%
61 to 70	21.29%	21.29%	18.75%	27.62%	19.26%
71 to 80	10.59%	14.47%	11.80%	24.22%	17.26%
81 to 90	5.57%	7.14%	5.27%	19.07%	10.44%
91 to 100	3.88%	6.71%	3.71%	16.07%	4.00%
MAPE	25.71%	26.53%	**21.98%**	25.88%	27.18%

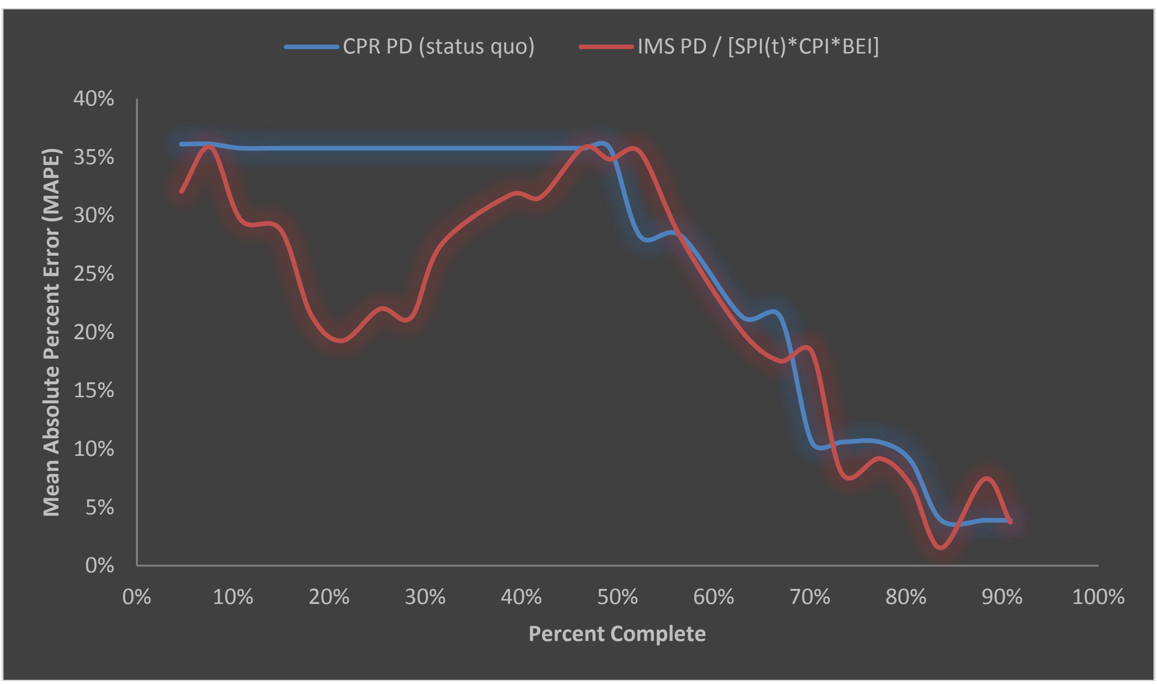


Figure 47: GPS OCX (FA8807-08-C-0003) Accuracy over Time

Table 67: WGS (FA8808-06-C-0001) Accuracy Results

Percent Complete Interval	Forecasting Model							
	CPR PD (status quo)	IMS PD	IMS PD / [SPI* CPI]	IMS PD / [SPI(t) (T.S.) * CPI]	Regress-ion	Kalman Filter	IDE	IDE / [SPI(t) (T.S.) * CPI]
0 to 10	44.55%	44.55%	46.56%	56.54%	23.32%	43.97%	44.55%	56.54%
11 to 20	43.44%	43.44%	41.37%	38.03%	14.65%	43.68%	43.44%	38.03%
21 to 30	43.66%	42.09%	39.54%	39.77%	33.84%	41.53%	39.96%	37.53%
31 to 40	29.95%	34.17%	31.39%	31.09%	36.60%	33.38%	18.58%	14.78%
41 to 50	16.58%	24.53%	22.79%	22.38%	32.36%	23.97%	27.94%	25.90%
51 to 60	21.10%	24.08%	22.56%	22.22%	28.56%	23.89%	18.77%	17.52%
61 to 70	16.75%	15.54%	12.18%	12.62%	25.31%	16.04%	9.82%	12.99%
71 to 80	16.75%	9.84%	7.11%	6.88%	22.52%	13.15%	8.59%	10.58%
81 to 90	16.75%	2.96%	1.05%	1.42%	17.49%	12.01%	2.96%	1.42%
91 to 100	16.75%	0.17%	2.62%	4.92%	2.35%	4.62%	0.17%	4.92%
MAPE	24.77%	22.03%	20.31%	20.31%	23.75%	23.70%	19.22%	**18.69%**

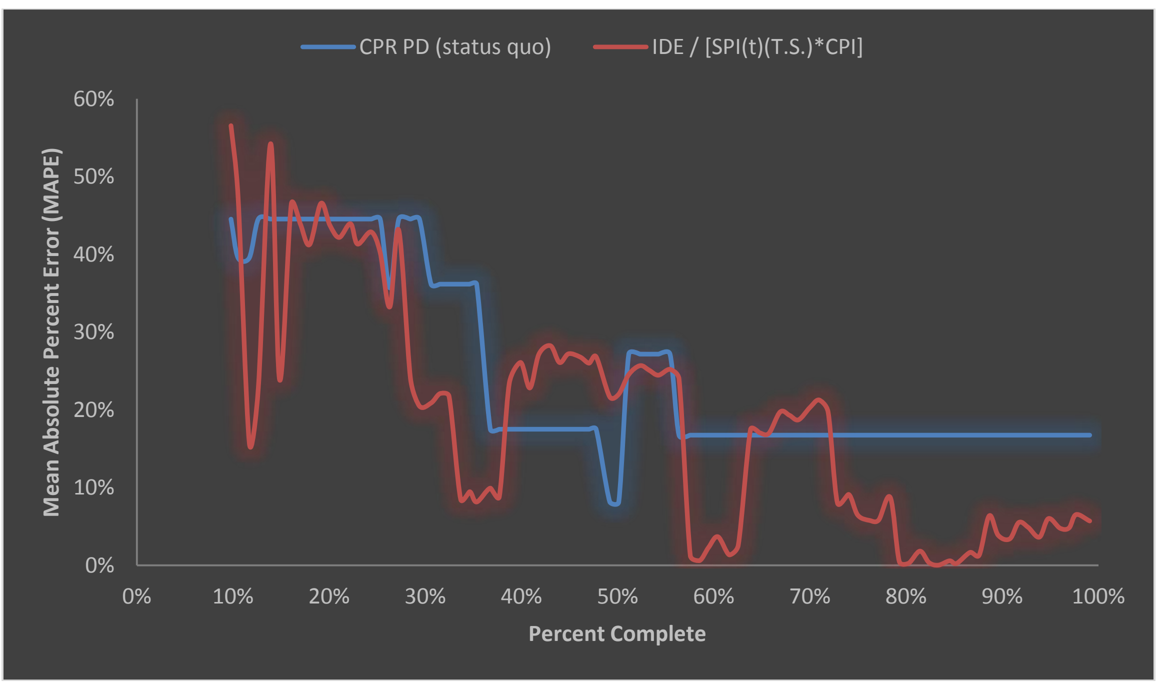


Figure 48: WGS (FA8808-06-C-0001) Accuracy over Time

Table 68: WGS (FA8808-10-C-0001) Accuracy Results

Percent Complete Interval	Forecasting Model						
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) (T.S.)* BEI]	Regression	Kalman Filter	IDE	IDE / [SPI(t) (T.S.) *BEI]
0 to 10	36.71%	37.16%	50.90%	20.46%	36.92%	37.16%	50.91%
11 to 20	36.73%	37.86%	51.18%	46.46%	39.38%	33.44%	47.71%
21 to 30	36.73%	37.07%	45.94%	53.88%	42.85%	27.82%	37.93%
31 to 40	36.73%	36.73%	40.47%	55.18%	41.40%	26.18%	30.53%
41 to 50	36.73%	36.89%	31.69%	51.94%	35.58%	21.70%	15.13%
51 to 60	36.73%	36.73%	29.95%	47.97%	33.03%	13.79%	4.58%
61 to 70	25.28%	25.28%	18.57%	43.70%	20.27%	9.01%	1.14%
71 to 80	20.97%	21.01%	17.61%	38.73%	16.29%	9.01%	5.44%
81 to 90	12.67%	12.67%	11.28%	32.71%	13.44%	10.75%	9.30%
91 to 100	0.00%	3.41%	3.44%	28.31%	4.22%	3.41%	3.44%
MAPE	29.33%	29.70%	30.90%	43.56%	29.47%	**19.53%**	20.45%

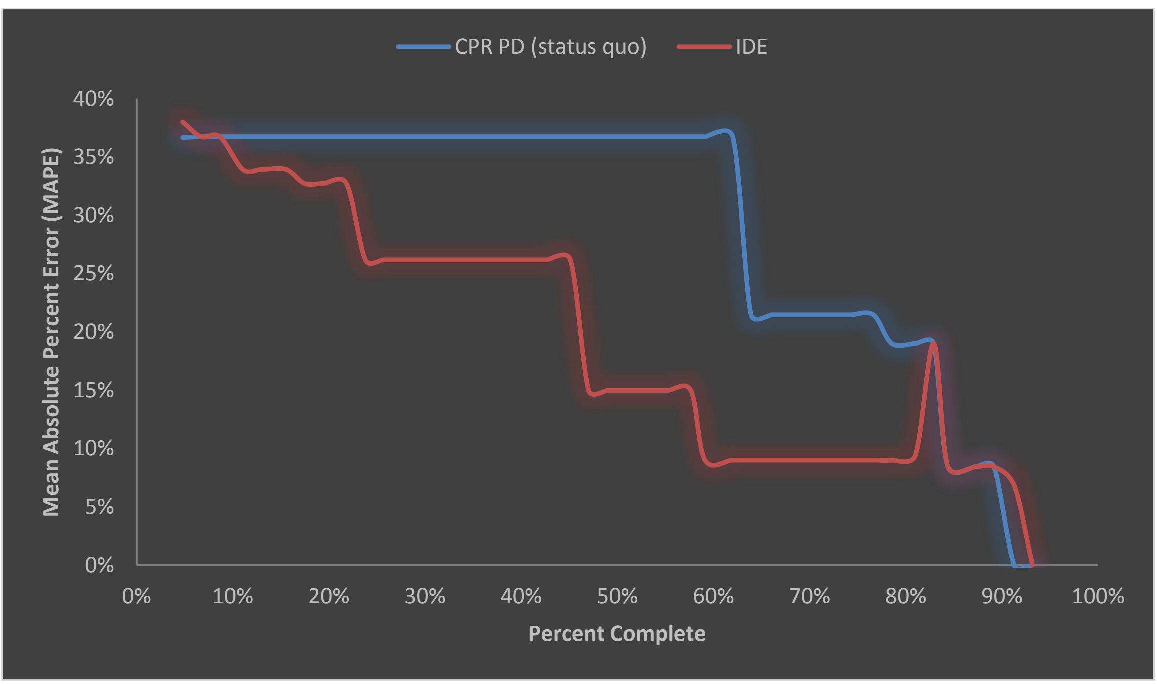


Figure 49: WGS (FA8808-10-C-0001) Accuracy Results

Table 69: MUOS (N00039-04-C-2009) Accuracy Results

Percent Complete Interval	Forecasting Model						
	CPR PD (status quo)	IMS PD	IMS PD/ SPI(T.S.)	Regression	Kalman Filter	IDE	IDE/ SPI(T.S.)
0 to 10	31.96%	28.35%	27.70%	11.07%	30.24%	28.09%	27.36%
11 to 20	31.96%	24.75%	23.89%	20.67%	30.10%	15.69%	14.75%
21 to 30	31.96%	22.90%	22.28%	26.06%	29.95%	10.99%	10.24%
31 to 40	31.96%	23.27%	21.81%	33.33%	29.21%	7.96%	8.49%
41 to 50	25.92%	22.24%	21.89%	31.48%	23.88%	2.07%	2.23%
51 to 60	16.36%	19.42%	18.55%	30.32%	7.02%	7.24%	7.65%
61 to 70	15.05%	3.40%	3.17%	29.13%	0.94%	3.27%	3.12%
71 to 80	9.82%	2.71%	2.23%	25.74%	0.89%	4.41%	4.27%
81 to 90	3.07%	1.75%	1.91%	21.23%	0.89%	7.11%	7.39%
91 to 100	2.03%	1.91%	2.03%	15.94%	3.79%	0.01%	0.12%
MAPE	19.23%	14.47%	13.97%	24.81%	14.92%	7.96%	**7.87%**

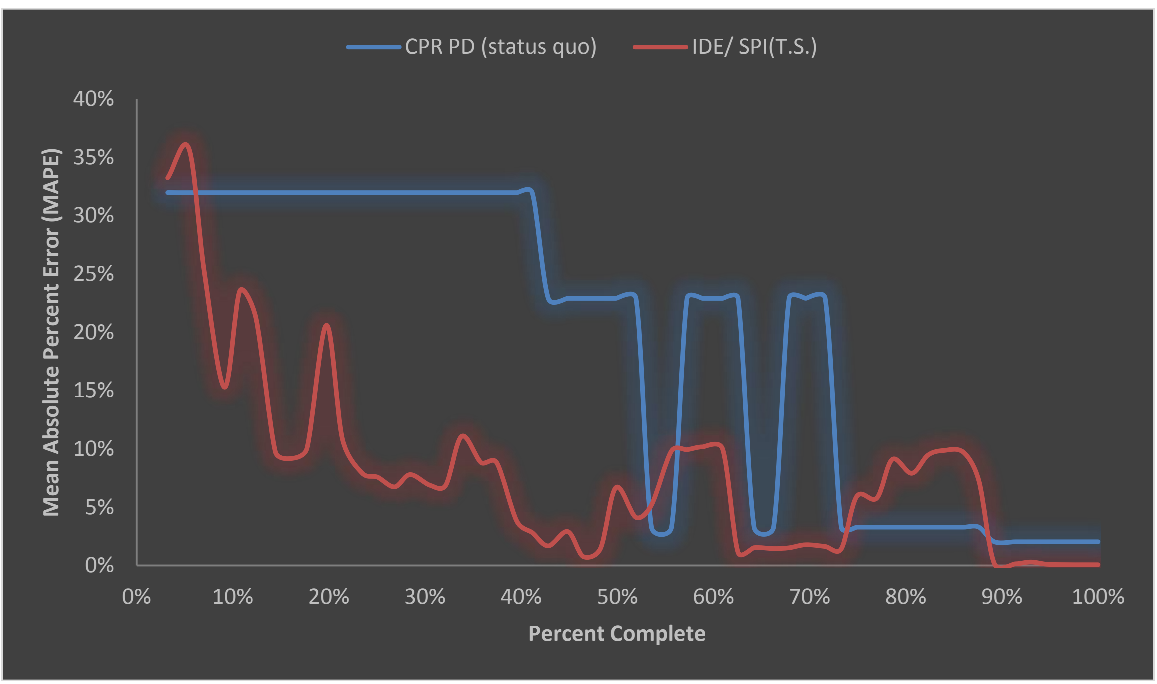


Figure 50: MUOS (N00039-04-C-2009) Accuracy over Time

Table 70: AEHF (F04701-02-C-0002) Accuracy Results

Percent Complete Interval	Forecasting Model				
	CPR PD (status quo)	IMS PD	IMS PD / [SPI(t) *CPI]	Regression	Kalman Filter
0 to 10	47.40%	47.40%	46.64%	60.90%	47.25%
11 to 20	46.87%	46.87%	44.51%	58.22%	46.25%
21 to 30	40.90%	40.90%	36.14%	37.23%	39.96%
31 to 40	30.92%	30.92%	23.53%	43.52%	29.73%
41 to 50	30.61%	30.61%	21.55%	44.35%	29.52%
51 to 60	24.66%	24.66%	19.68%	30.87%	25.94%
61 to 70	16.50%	16.50%	18.42%	24.97%	22.65%
71 to 80	13.19%	13.19%	15.16%	12.18%	13.17%
81 to 90	6.02%	6.02%	6.56%	5.37%	7.98%
91 to 100	2.36%	2.36%	5.93%	6.03%	2.98%
MAPE	25.66%	25.66%	**23.09%**	31.72%	26.32%

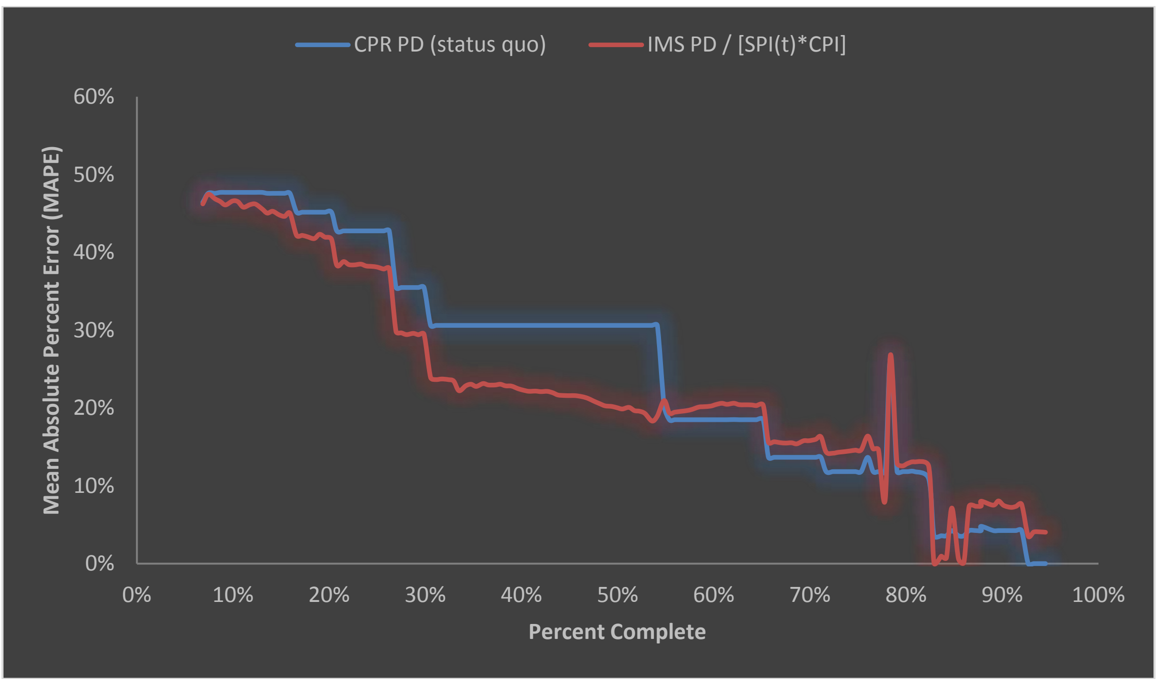


Figure 51: AEHF (F04701-02-C-0002) Accuracy Results over Time

Table 71: SBIRS (F04701-95-C-0017) Accuracy Results

Percent Complete Interval	Forecasting Model						
	CPR PD (status quo)	IMS PD	IDE	IMS PD / [SPI(t) (T.S.) *BEI (T.S.)]	IDE/ SPI	Regression	Kalman Filter
0 to 10	50.61%	50.61%	50.61%	49.98%	50.26%	57.67%	50.34%
11 to 20	46.45%	46.45%	46.45%	45.83%	45.78%	63.40%	45.96%
21 to 30	37.54%	37.54%	37.54%	36.88%	36.69%	60.71%	36.61%
31 to 40	31.33%	31.33%	31.33%	26.14%	30.60%	39.21%	30.55%
41 to 50	24.85%	24.85%	24.85%	13.54%	24.60%	29.83%	24.93%
51 to 60	16.83%	16.89%	16.19%	5.61%	16.07%	30.62%	17.21%
61 to 70	11.04%	10.67%	3.10%	1.75%	3.87%	20.49%	10.87%
71 to 80	2.50%	4.85%	3.38%	6.88%	3.66%	8.17%	3.94%
81 to 90	0.14%	0.04%	7.94%	10.71%	8.07%	8.93%	0.27%
91 to 100							
MAPE	24.63%	24.84%	24.60%	**21.88%**	24.40%	35.60%	24.56%

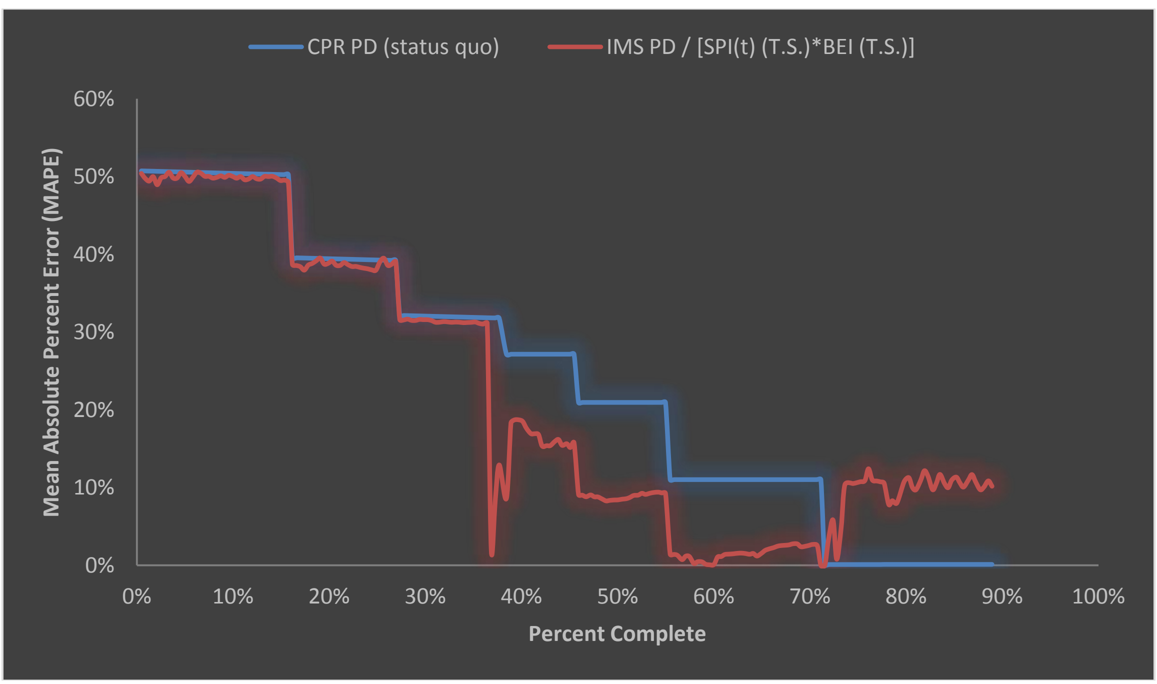
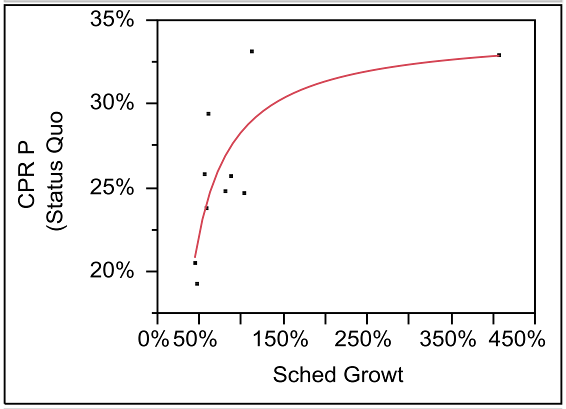


Figure 52: SBIRS (F04701-95-C-0017) Accuracy Results over Time

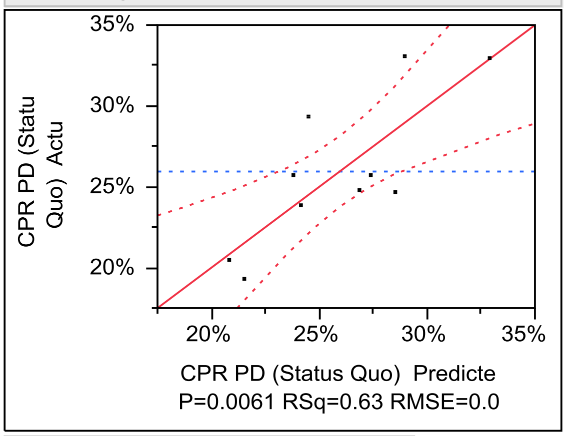
Appendix D: Regression Analysis Outputs

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.629764
RSquare Adj	0.583485
Root Mean Square Error	0.029958
Mean of Response	0.25944
Observations (or Sum Wgts	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.01221240	0.012212	13.6078
Error	8	0.00717962	0.000897	**Prob > F**
C. Tota	9	0.01939202		0.0061 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.3436491	0.024715	13.90	<.0001 *	0
Reciprocal(Sched Growth	-0.06037	0.016365	-3.69	0.0061 *	-0.79358

Figure 53: Regression Output - CPR PD (status quo)

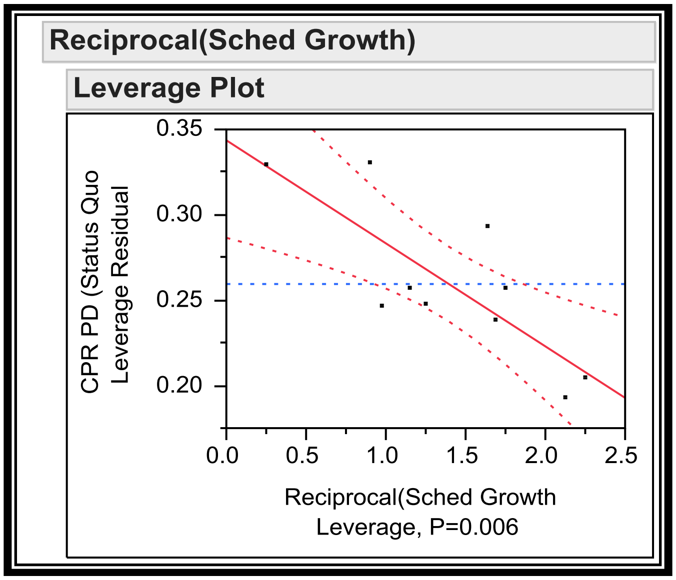


Figure 54: Leverage Plot - CPR PD (status quo)

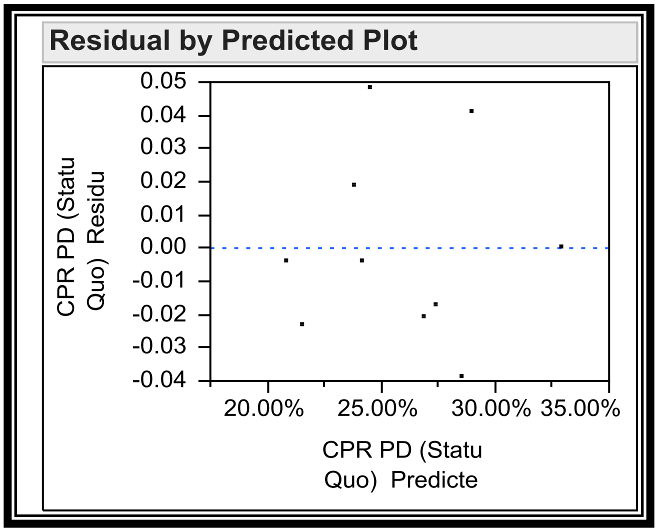


Figure 55: Residual Plot - CPR PD (status quo)

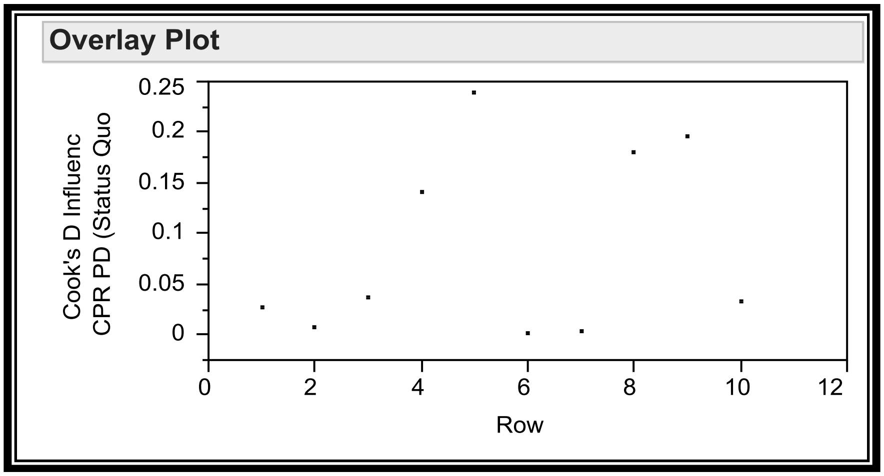


Figure 56: Cook's D - CPR PD (status quo)

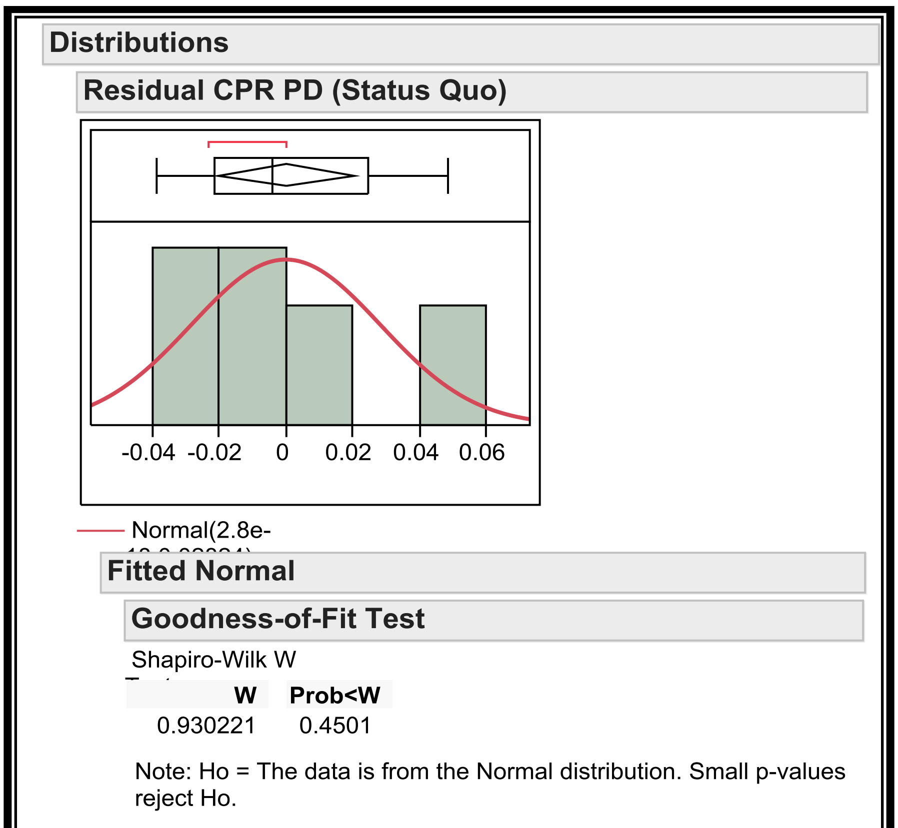
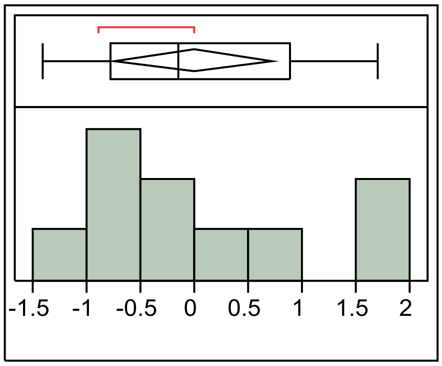


Figure 57: Residuals Histogram & Shapiro-Wilk Normality Test (CPR PD)

Distributions

Studentized Resid CPR PD (Status Quo)



Quantiles

100.0	maximu	1.71413
99.5%		1.71413
97.5%		1.71413
90.0%		1.69343
75.0%	quartile	0.88753
50.0%	median	-0.1552
25.0%	quartile	-0.7758
10.0%		-1.3554
2.5%		-1.4061
0.5%		-1.4061
0.0%	minimum	-1.4061

Summary Statistics

Mean	-0.006231
Std Dev	1.023461
Std Err Mean	0.3236468
Upper 95% Mea	0.7259084
Lower 95% Mea	-0.738371
N	10

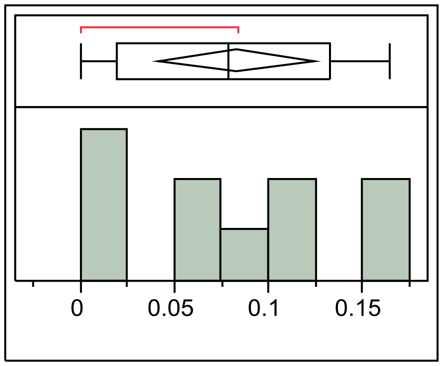
Figure 58: Studentized Residuals Check for Outliers (CPR PD)

Table 72: Breusch-Pagan Test for Heteroscedasticity (CPR PD)

N	10
Degrees of Freedom model	1
Sum of Squared Errors (SSE)	0.007180
Sum of Squared Residuals (SSR)	8.09E-08
Breusch-Pagan Test Statistic	0.0784
Breusch-Pagan Test p-value	0.7794

Distributions

APE



Quantiles

100.0	maximu	0.16452
99.5%		0.16452
97.5%		0.16452
90.0%		0.1638
75.0%	quartile	0.13246
50.0%	median	0.07885
25.0%	quartile	0.01887
10.0%		0.00197
2.5%		0.00019
0.5%		0.00019
0.0%	minimum	0.00019

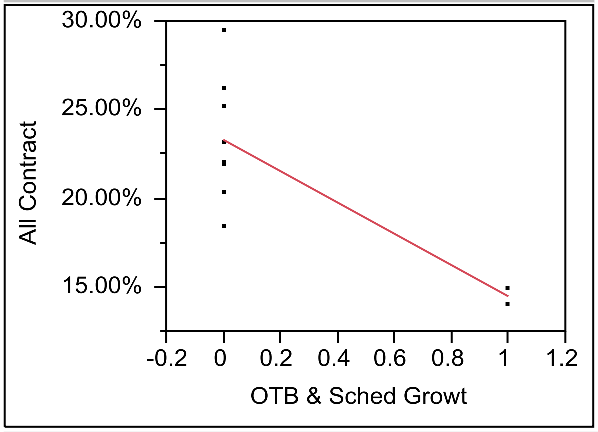
Summary Statistics

Mean	0.0829375
Std Dev	0.0584632
Std Err Mean	0.0184877
Upper 95% Mea	0.1247595
Lower 95% Mea	0.0411155
N	10

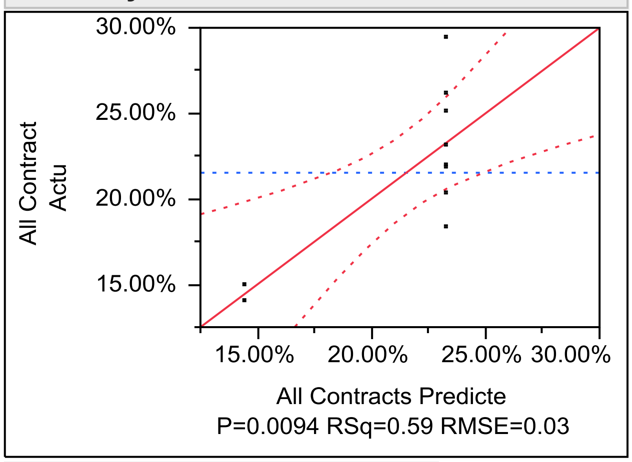
Figure 59: MAPE - CPR PD (status quo)

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.590317
RSquare Adj	0.539106
Root Mean Square Error	0.032981
Mean of Response	0.21527
Observations (or Sum Wgts	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.01253868	0.012539	11.5273
Error	8	0.00870192	0.001088	**Prob > F**
C. Tota	9	0.02124060		0.0094 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.232975	0.011661	19.98	<.0001 *	0
OTB & Sched Growt	-0.088525	0.026074	-3.40	0.0094 *	-0.76832

Figure 60: Regression Output (IMS MAPEs)

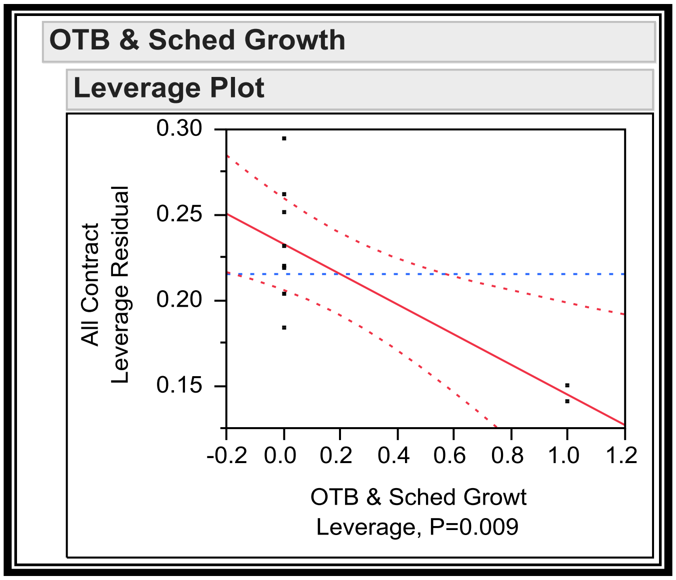


Figure 61: Leverage Plot (IMS MAPEs)

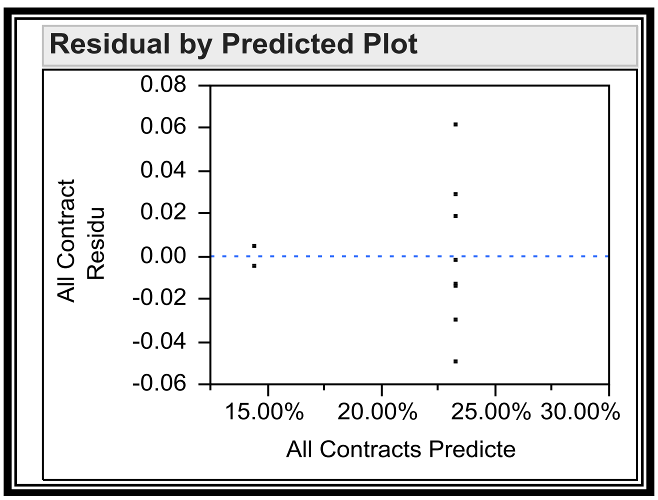


Figure 62: Residuals Plot (IMS MAPEs)

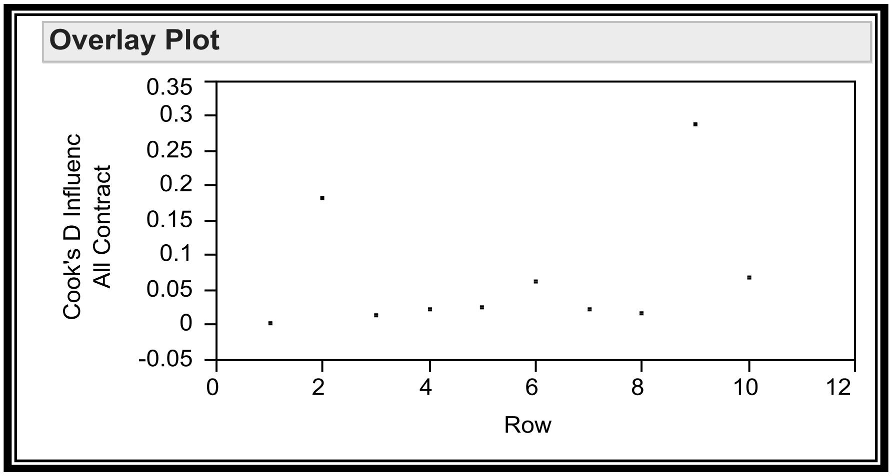


Figure 63: Cook's D (IMS Model MAPEs)

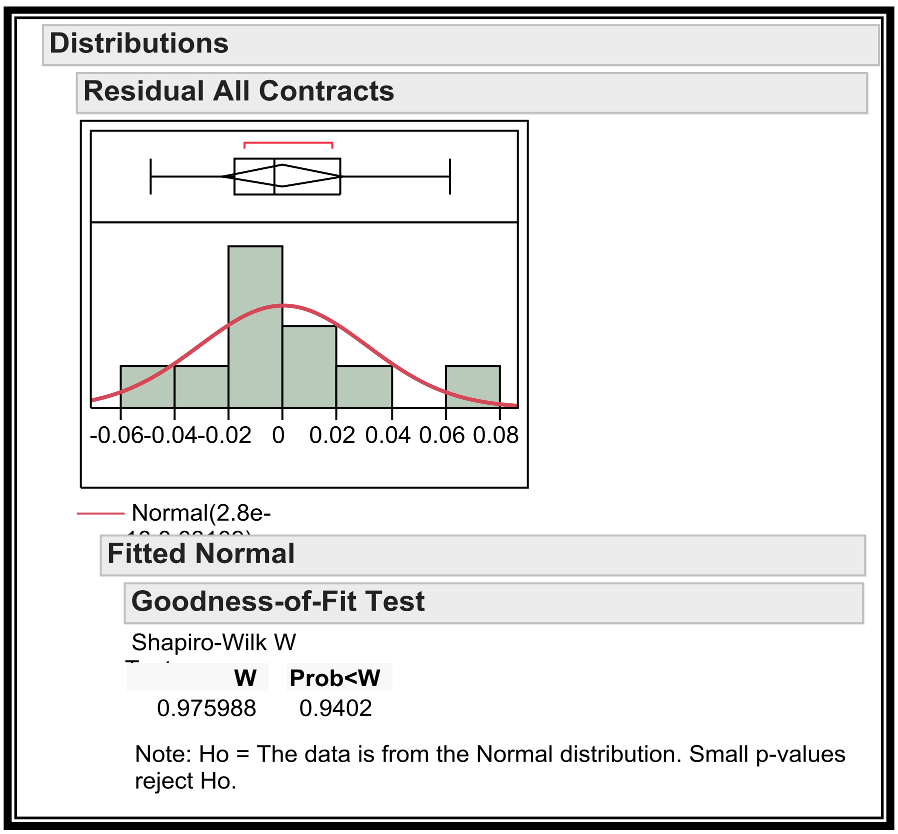
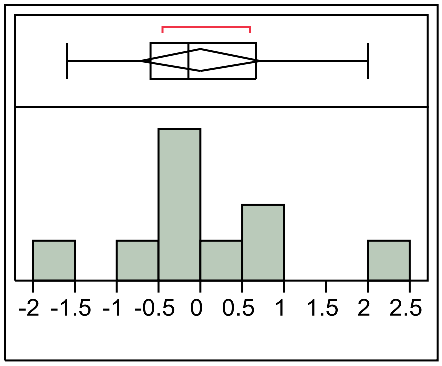


Figure 64: Residuals Histogram & Shapiro-Wilk Normality Test (IMS MAPEs)

Distributions

Studentized Residual All Contracts



Quantiles

100.0	maximu	2.00076
99.5%		2.00076
97.5%		2.00076
90.0%		1.89282
75.0%	quartile	0.67826
50.0%	median	-0.1355
25.0%	quartile	-0.5867
10.0%		-1.5343
2.5%		-1.5972
0.5%		-1.5972
0.0%	minimum	-1.5972

Summary Statistics

Mean	4.441e-17
Std Dev	1.0098633
Std Err Mean	0.3193468
Upper 95% Mea	0.7224127
Lower 95% Mea	-0.722413
N	10

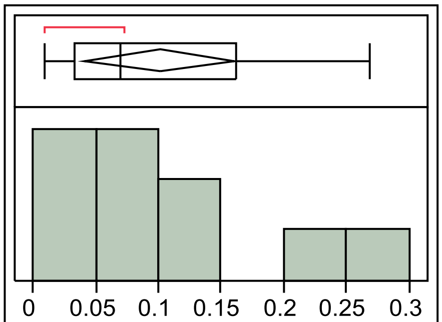
Figure 65: Studentized Residuals Check for Outliers (IMS MAPEs)

Table 73: Breusch-Pagan Test for Heteroscedasticity (IMS MAPEs)

N	10
Degrees of Freedom model	1
Sum of Squared Errors (SSE)	0.008702
Sum of Squared Residuals (SSR)	1.80E-06
Breusch-Pagan Test Statistic	1.1885
Breusch-Pagan Test p-value	0.2756

Distributions

APE



Quantiles

100.0	maximu	0.26824
99.5%		0.26824
97.5%		0.26824
90.0%		0.26236
75.0%	quartile	0.16268
50.0%	median	0.06904
25.0%	quartile	0.03346
10.0%		0.01127
2.5%		0.00899
0.5%		0.00899
0.0%	minimum	0.00899

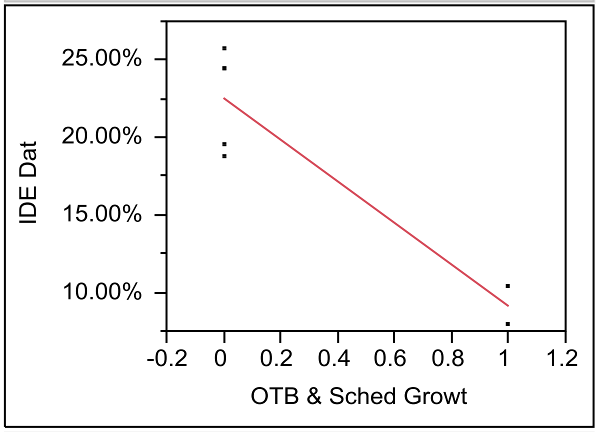
Summary Statistics

Mean	0.1006363
Std Dev	0.0839424
Std Err Mean	0.0265449
Upper 95% Mea	0.1606851
Lower 95% Mea	0.0405876
N	10

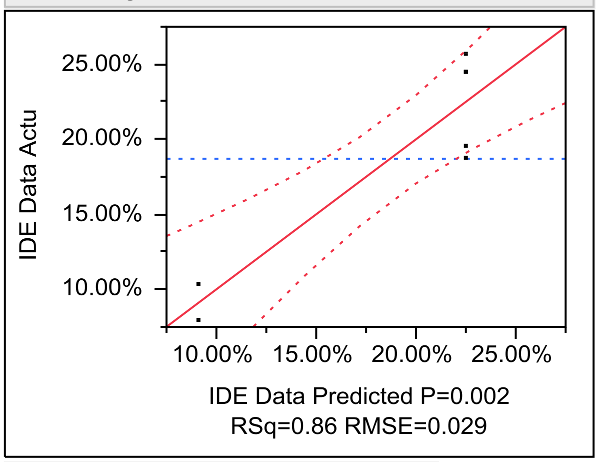
Figure 66: MAPE for Predicting IMS Model Accuracy

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.855006
RSquare Adj	0.826007
Root Mean Square Error	0.029602
Mean of Response	0.187057
Observations (or Sum Wgts	7

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.02583553	0.025836	29.4841
Error	5	0.00438127	0.000876	**Prob > F**
C. Tota	6	0.03021680		0.0029 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.22548	0.013238	17.03	<.0001 *	0
OTB & Sched Growt	-0.13448	0.024766	-5.43	0.0029 *	-0.92467

Figure 67: Regression Output (IDE MAPEs)

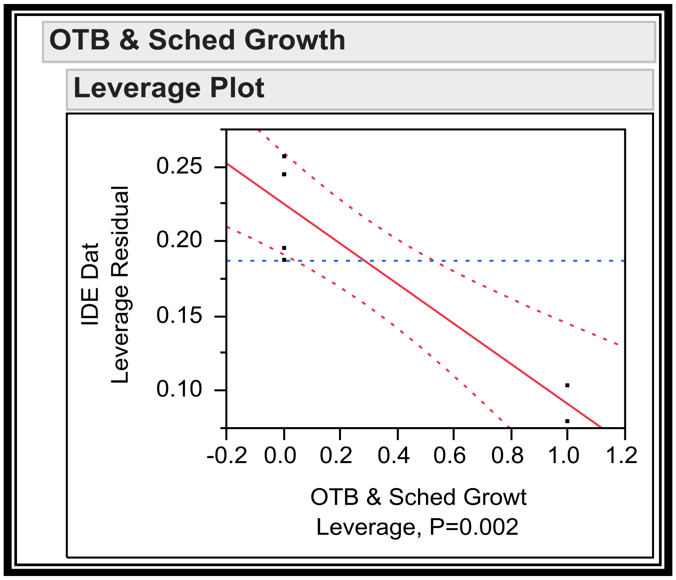


Figure 68: Leverage Plot (IDE MAPEs)

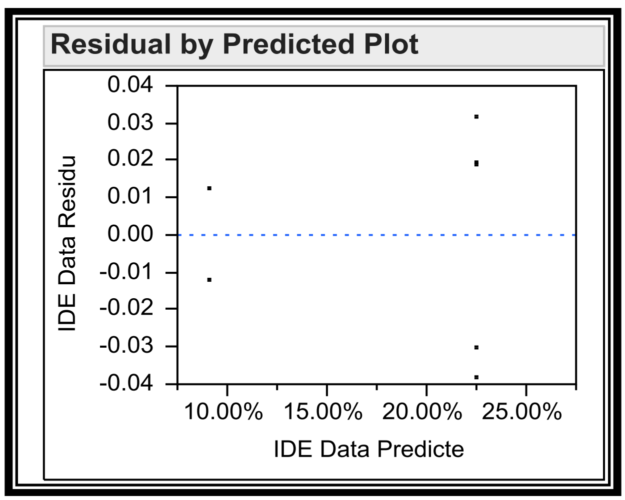


Figure 69: Residuals Plot (IDE MAPEs)

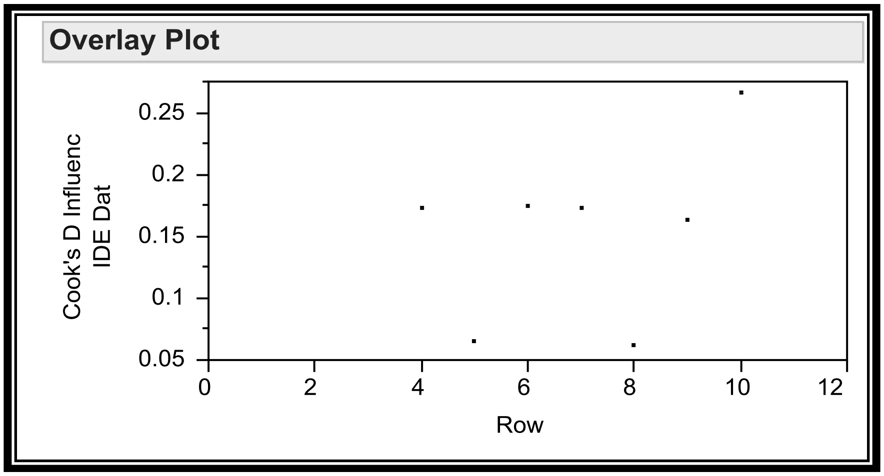


Figure 70: Cook's D (IDE MAPEs)

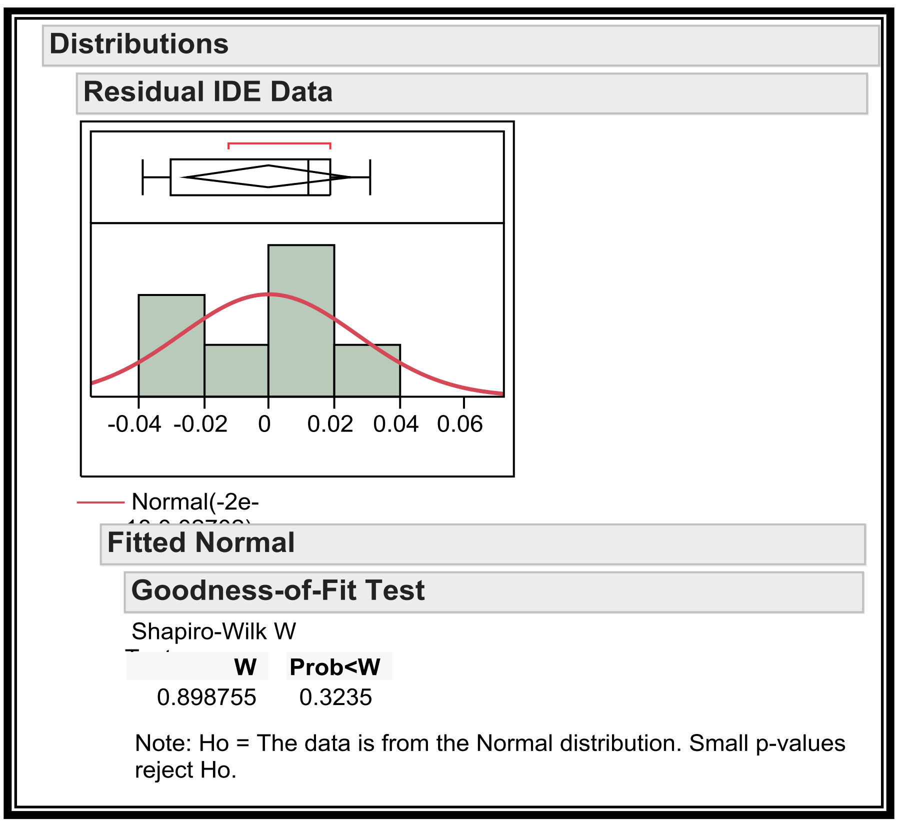
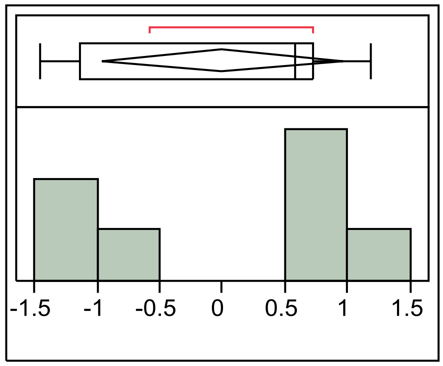


Figure 71: Residuals Histogram & Shapiro-Wilk Normality Test (IDE MAPEs)

Distributions

Studentized Resid IDE Data



Quantiles

100.0	maximu	1.17916
99.5%		1.17916
97.5%		1.17916
90.0%		1.17916
75.0%	quartile	0.71837
50.0%	median	0.58763
25.0%	quartile	-1.1399
10.0%		-1.4571
2.5%		-1.4571
0.5%		-1.4571
0.0%	minimum	-1.4571

Summary Statistics

Mean	-6.34e-17
Std Dev	1.041552
Std Err Mean	0.3936696
Upper 95% Mea	0.9632749
Lower 95% Mea	-0.963275
N	7

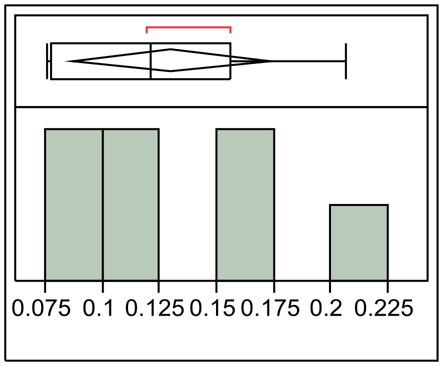
Figure 72: Studentized Residuals Check for Outliers (IDE MAPEs)

Table 74: Breusch-Pagan Test for Heteroscedasticity (IDE MAPEs)

N	7
Degrees of Freedom model	1
Sum of Squared Errors (SSE)	0.004381
Sum of Squared Residuals (SSR)	6.31E-07
Breusch-Pagan Test Statistic	0.8050
Breusch-Pagan Test p-value	0.3696

Distributions

APE



Quantiles

100.0	maximu	0.20642
99.5%		0.20642
97.5%		0.20642
90.0%		0.20642
75.0%	quartile	0.15629
50.0%	median	0.12162
25.0%	quartile	0.07779
10.0%		0.0759
2.5%		0.0759
0.5%		0.0759
0.0%	minimum	0.0759

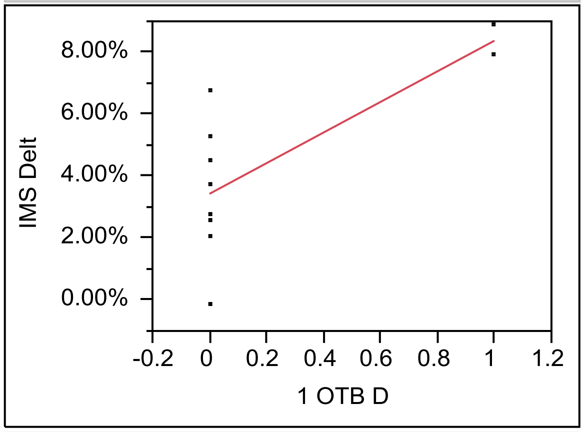
Summary Statistics

Mean	0.1302323
Std Dev	0.0465021
Std Err Mean	0.0175762
Upper 95% Mea	0.1732396
Lower 95% Mea	0.087225
N	7

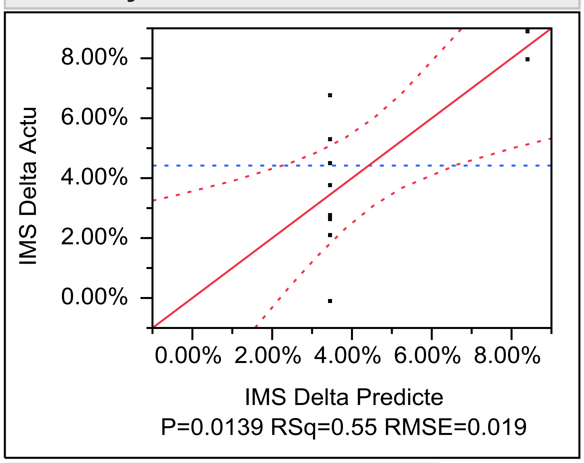
Figure 73: MAPE for Predicting IDE Model Accuracy

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.551606
RSquare Adj	0.495557
Root Mean Square Error	0.019944
Mean of Response	0.04418
Observations (or Sum Wgts	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.00391446	0.003914	9.8415
Error	8	0.00318201	0.000398	**Prob > F**
C. Tota	9	0.00709648		0.0139 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.0342875	0.007051	4.86	0.0013 *	0
1 OTB D	0.0494625	0.015767	3.14	0.0139 *	0.742702

Figure 74: Regression Output (IMS MAPE Delta)

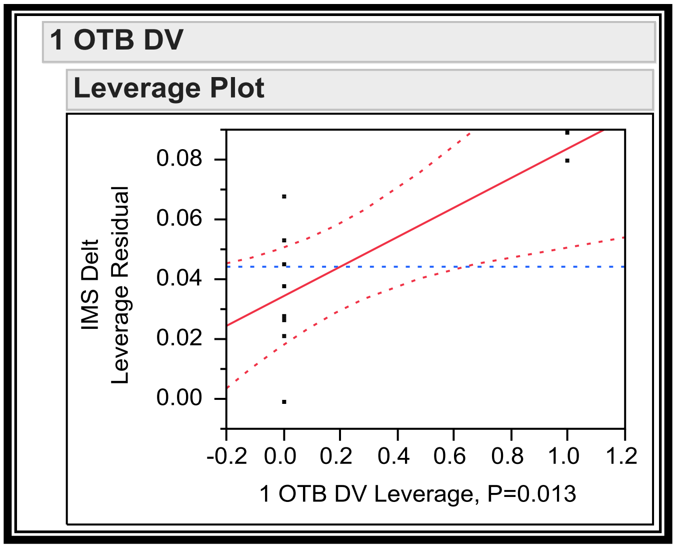


Figure 75: Leverage Plot (IMS MAPE Delta)

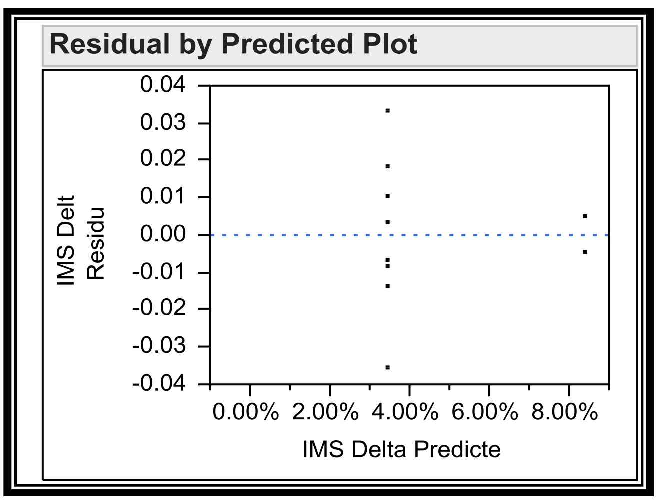


Figure 76: Residuals Plot (IMS MAPE Delta)

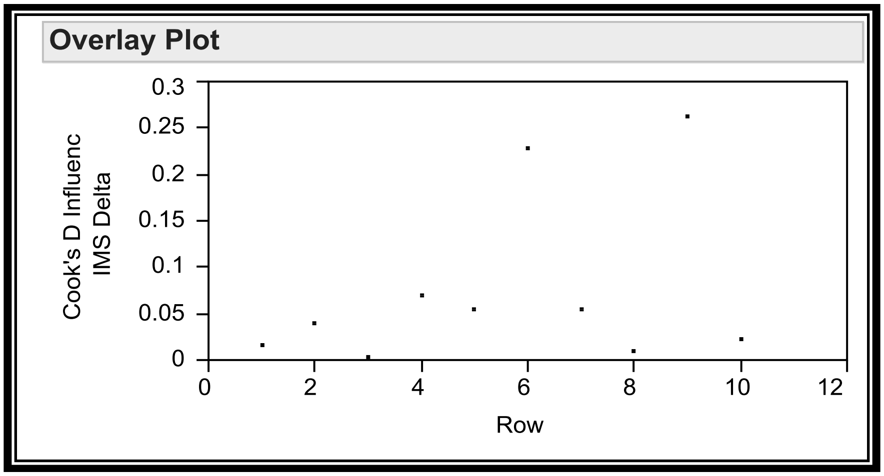


Figure 77: Cook's D (IMS MAPE Delta)

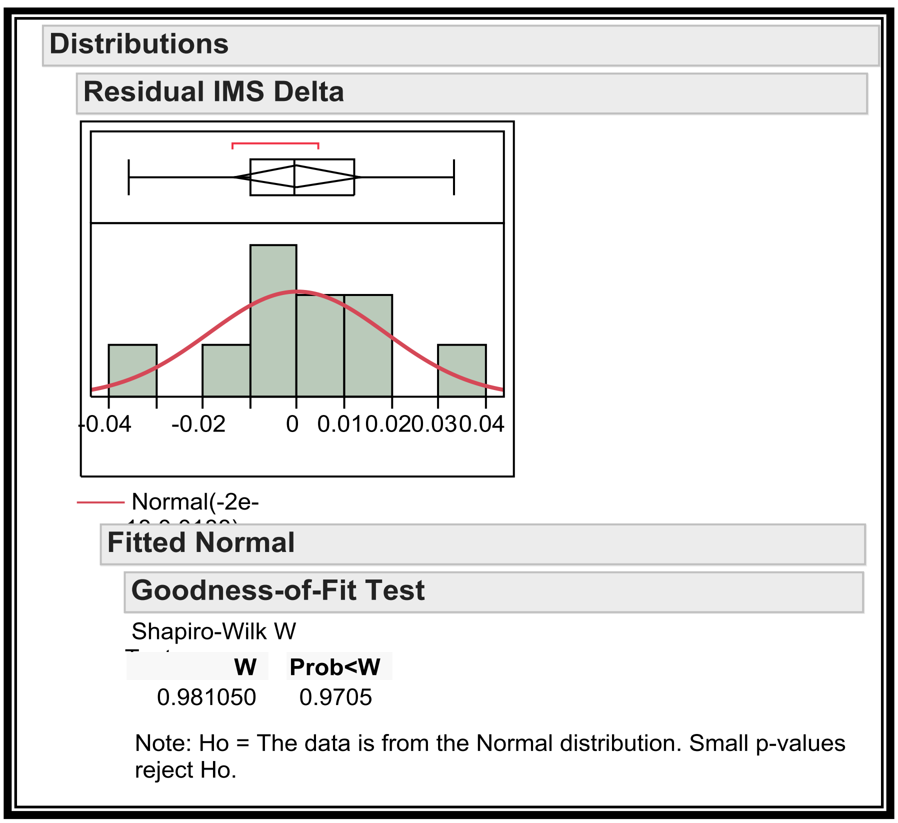


Figure 78: Residuals Histogram & Shapiro-Wilk Test (IMS MAPE Delta)

Distributions

Studentized Resid IMS Delta

Quantiles

100.0	maximu	1.78029
99.5%		1.78029
97.5%		1.78029
90.0%		1.70042
75.0%	quartile	0.65999
50.0%	median	-0.0841
25.0%	quartile	-0.5313
10.0%		-1.7961
2.5%		-1.913
0.5%		-1.913
0.0%	minimum	-1.913

Summary Statistics

Mean	-1.22e-16
Std Dev	1.0130289
Std Err Mean	0.3203479
Upper 95% Mea	0.7246773
Lower 95% Mea	-0.724677
N	10

Figure 79: Studentized Residuals Check for Outliers (IMS MAPE Delta)

Table 75: Breusch-Pagan Test for Heteroscedasticity (IMS MAPE Delta)

N	10
Degrees of Freedom model	1
Sum of Squared Errors (SSE)	0.003182
Sum of Squared Residuals (SSR)	2.20E-07
Breusch-Pagan Test Statistic	1.0859
Breusch-Pagan Test p-value	0.2974

Distributions

APE

Quantiles

100.0	maximu	25.4911
99.5%		25.4911
97.5%		25.4911
90.0%		23.01
75.0%	quartile	0.53922
50.0%	median	0.28822
25.0%	quartile	0.07527
10.0%		0.05322
2.5%		0.0526
0.5%		0.0526
0.0%	minimum	0.0526

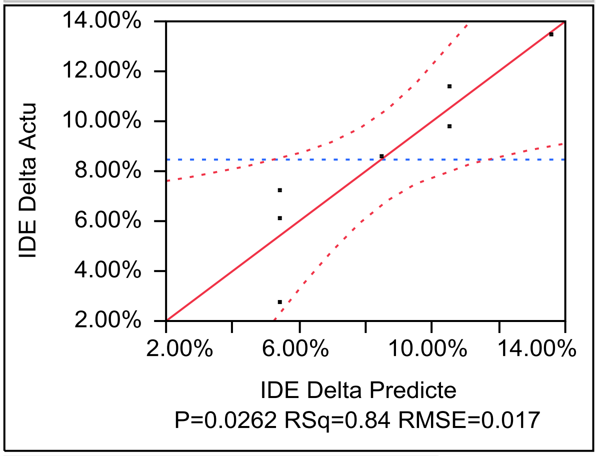
Summary Statistics

Mean	2.8011834
Std Dev	7.9749034
Std Err Mean	2.5218859
Upper 95% Mea	8.5060855
Lower 95% Mea	-2.903719
N	10

Figure 80: MAPE for Predicting the Accuracy Delta (IMS Models - CPR PD)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.838253
RSquare Adj	0.757379
Root Mean Square Error	0.017364
Mean of Response	0.084643
Observations (or Sum Wgts	7

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.00625056	0.003125	10.3650
Error	4	0.00120609	0.000302	**Prob > F**
C. Tota	6	0.00745666		0.0262 *

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fi	1	0.00000532	5.322e-6	0.0133
Pure Erro	3	0.00120077	0.000400	**Prob > F**
Total Erro	4	0.00120609		0.9155
				Max RSq
				0.8390

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta	VIF
Intercept	0.0540235	0.009417	5.74	0.0046 *	0	.
Sched Growth <.6	0.0510412	0.013318	3.83	0.0186 *	0.77391	1.0084034
1 OTB DV	0.0306059	0.014589	2.10	0.1039	0.423627	1.0084034

Figure 81: Regression Output #1 (IDE MAPE Delta)

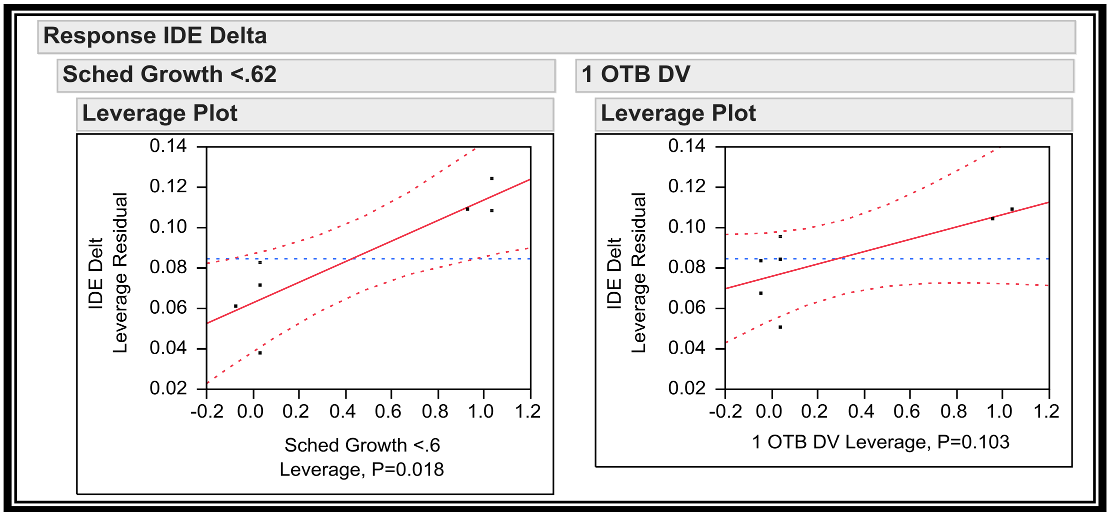


Figure 82: Leverage Plots (IDE MAPE Delta)

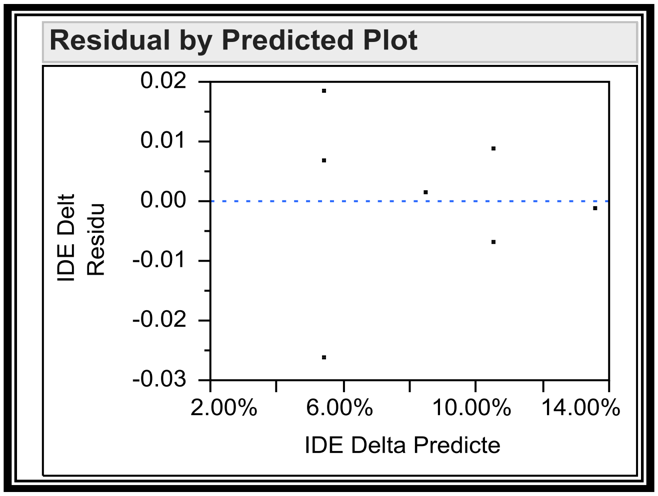


Figure 83: Residuals Plot (IDE MAPE Delta)

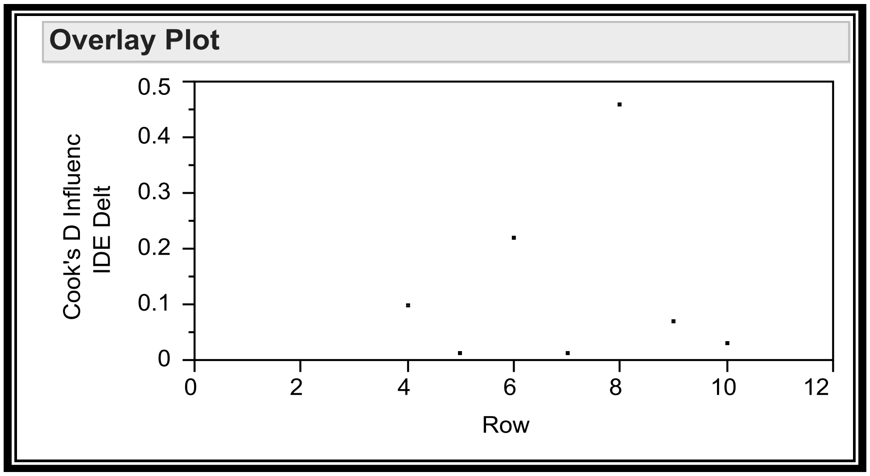


Figure 84: Cook's D (IDE MAPE Delta)

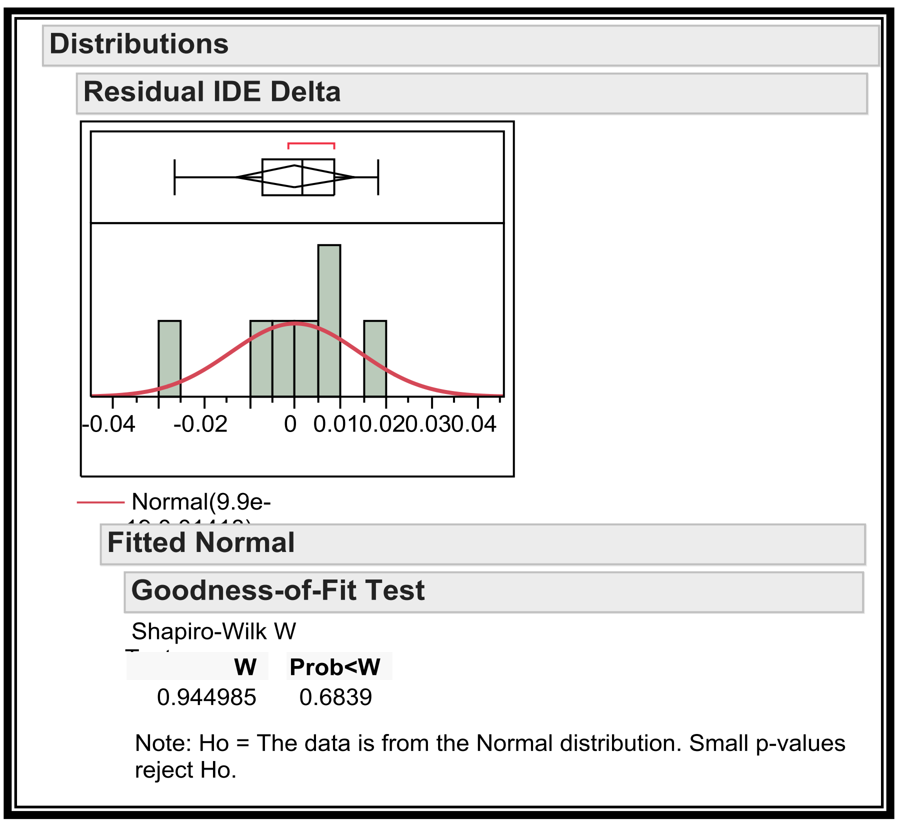


Figure 85: Residuals Histogram & Shapiro-Wilk Test (IDE MAPE Delta)

Distributions

Studentized Resid IDE Delta

Quantiles

100.0	maximu	1.25275
99.5%		1.25275
97.5%		1.25275
90.0%		1.25275
75.0%	quartile	0.64089
50.0%	median	0.13286
25.0%	quartile	-0.538
10.0%		-1.8112
2.5%		-1.8112
0.5%		-1.8112
0.0%	minimum	-1.8112

Summary Statistics

Mean	0.001281
Std Dev	0.9832768
Std Err Mean	0.3716437
Upper 95% Mea	0.9106603
Lower 95% Mea	-0.908098
N	7

Figure 86: Studentized Residuals Check for Outliers (IDE MAPE Delta)

Table 76: Breusch-Pagan Test for Heteroscedasticity (IDE MAPE Delta)

N	7
Degrees of Freedom model	2
Sum of Squared Errors (SSE)	0.001206
Sum of Squared Residuals (SSR)	3.12E-08
Breusch-Pagan Test Statistic	0.5254
Breusch-Pagan Test p-value	0.7690

Distributions

APE

Quantiles

100.0	maximu	0.95737
99.5%		0.95737
97.5%		0.95737
90.0%		0.95737
75.0%	quartile	0.25279
50.0%	median	0.07513
25.0%	quartile	0.01594
10.0%		0.01021
2.5%		0.01021
0.5%		0.01021
0.0%	minimum	0.01021

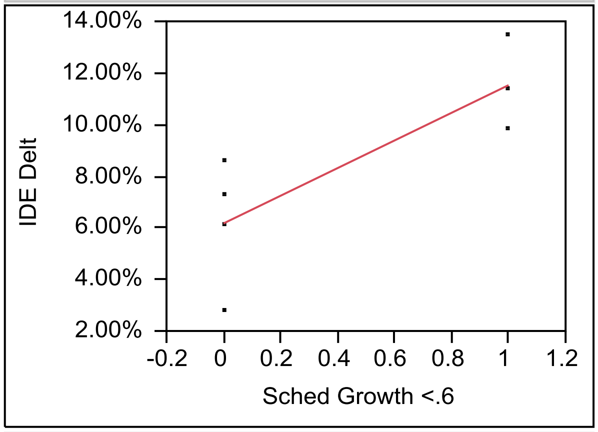
Summary Statistics

Mean	0.2137253
Std Dev	0.3377619
Std Err Mean	0.127662
Upper 95% Mea	0.5261029
Lower 95% Mea	-0.098652
N	7

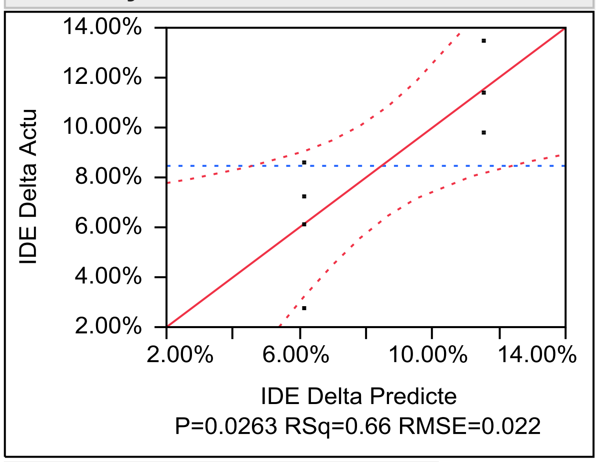
Figure 87: MAPE for Predicting the Accuracy Delta (IDE - CPR PD)

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.660288
RSquare Adj	0.592346
Root Mean Square Error	0.022508
Mean of Response	0.084643
Observations (or Sum Wgts	7

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.00492354	0.004924	9.7184
Error	5	0.00253311	0.000507	**Prob > F**
C. Tota	6	0.00745666		0.0263 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.061675	0.011254	5.48	0.0028 *	0
Sched Growth <.6	0.0535917	0.017191	3.12	0.0263 *	0.812581

Figure 88: Regression Output #2 (IDE MAPE Delta)

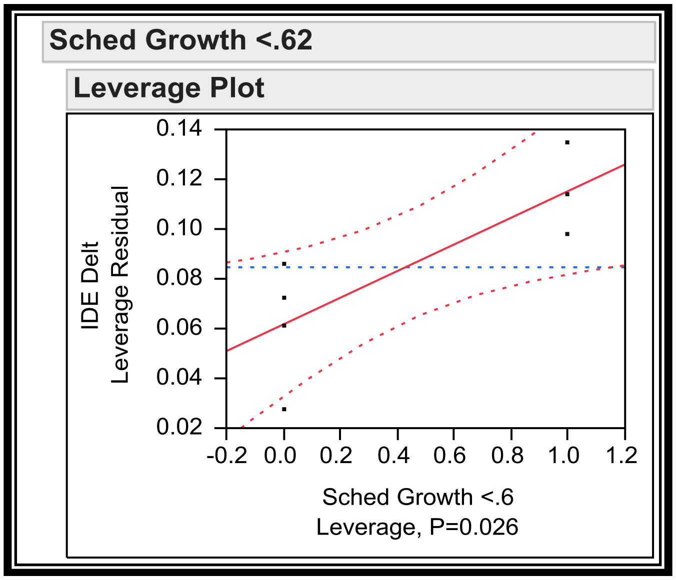


Figure 89: Leverage Plot (IDE MAPE Delta)

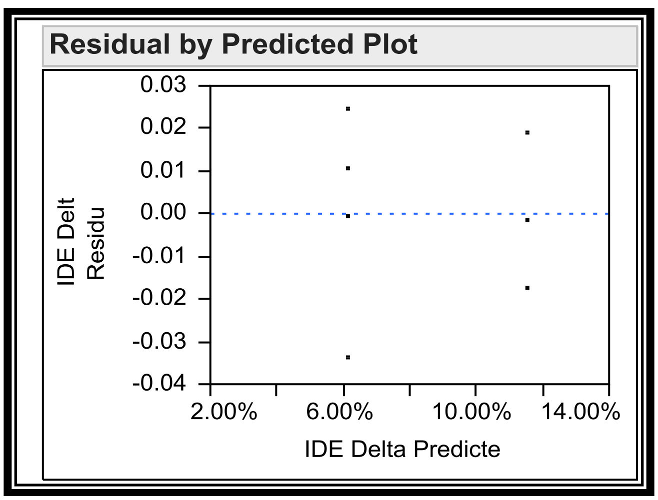


Figure 90: Residuals Plot (IDE MAPE Delta)

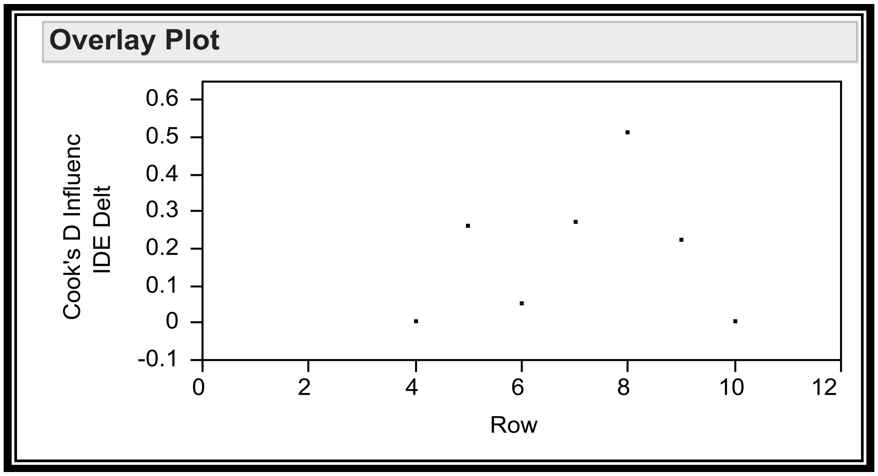


Figure 91: Cook's D (IDE MAPE Delta)

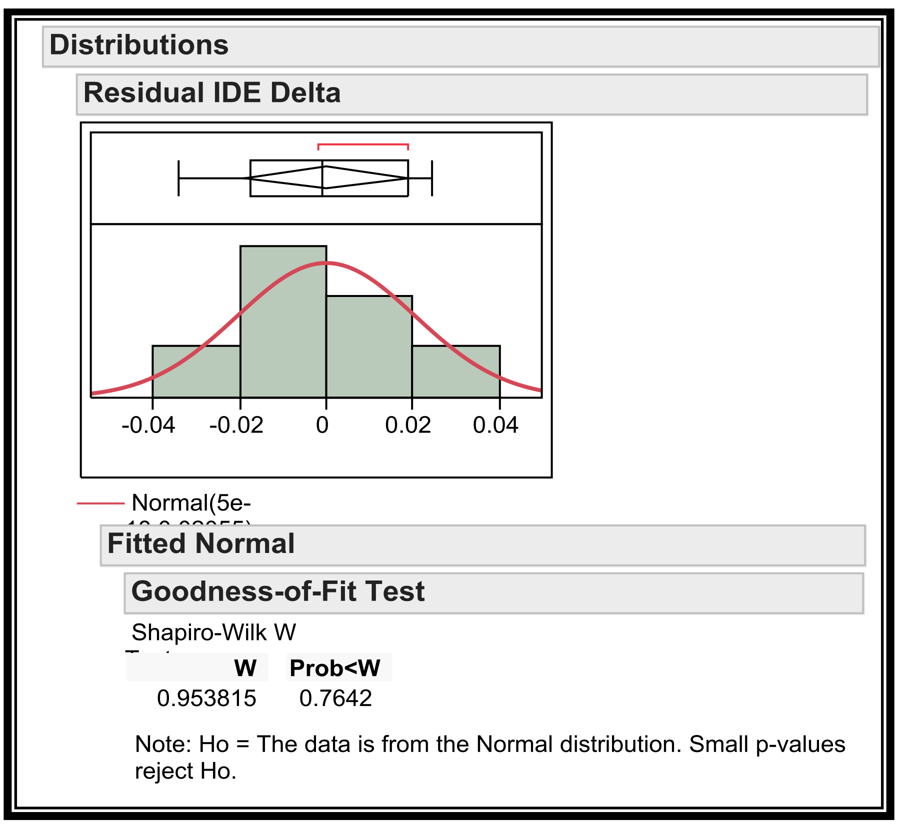


Figure 92: Residuals Histogram & Shapiro-Wilk Test (IDE MAPE Delta)

Distributions

Studentized Resid IDE Delta

Quantiles

100.0	maximu	1.2479
99.5%		1.2479
97.5%		1.2479
90.0%		1.2479
75.0%	quartile	1.03566
50.0%	median	-0.0449
25.0%	quartile	-0.945
10.0%		-1.7481
2.5%		-1.7481
0.5%		-1.7481
0.0%	minimum	-1.7481

Summary Statistics

Mean	2.548e-16
Std Dev	1.0712903
Std Err Mean	0.4049097
Upper 95% Mea	0.9907783
Lower 95% Mea	-0.990778
N	7

Figure 93: Studentized Residuals Check for Outliers (IDE MAPE Delta)

Table 77: Breusch-Pagan Test for Heteroscedasticity (IDE MAPE Delta)

N	7
Degrees of Freedom model	1
Sum of Squared Errors (SSE)	0.002533
Sum of Squared Residuals (SSR)	1.02E-07
Breusch-Pagan Test Statistic	0.3910
Breusch-Pagan Test p-value	0.5318

Distributions

APE

Quantiles

100.0	maximu	1.2346
99.5%		1.2346
97.5%		1.2346
90.0%		1.2346
75.0%	quartile	0.28285
50.0%	median	0.14696
25.0%	quartile	0.01467
10.0%		0.01439
2.5%		0.01439
0.5%		0.01439
0.0%	minimum	0.01439

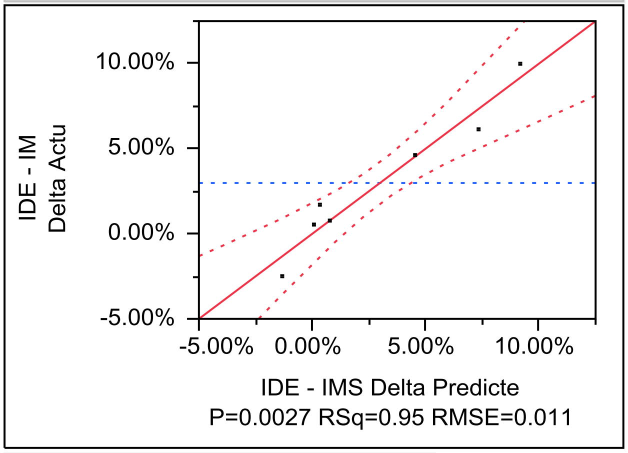
Summary Statistics

Mean	0.2875121
Std Dev	0.4280136
Std Err Mean	0.1617739
Upper 95% Mea	0.6833586
Lower 95% Mea	-0.108334
N	7

Figure 94: MAPE for Predicting the Accuracy Delta (IDE - CPR PD)

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.948416
RSquare Adj	0.922624
Root Mean Square Error	0.011613
Mean of Response	0.029843
Observations (or Sum Wgts	7

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.00991895	0.004959	36.7718
Error	4	0.00053949	0.000135	**Prob > F**
C. Tota	6	0.01045844		0.0027 *

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta	VIF
Intercept	0.0135335	0.005589	2.42	0.0727	0	.
Log(Cost Growth	-0.025682	0.003	-8.56	0.0010 *	-1.03864	1.1412582
1 OTB DV	0.0264452	0.01038	2.55	0.0635	0.309075	1.1412582

Residual by Predicted Plot

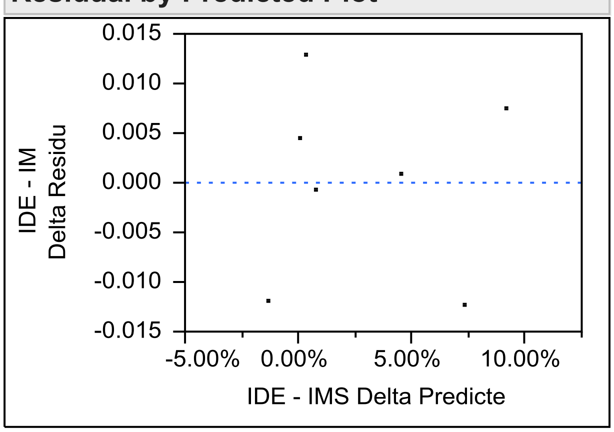
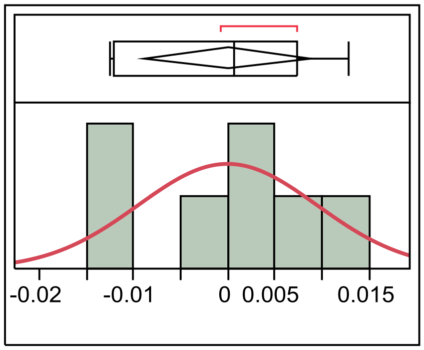


Figure 95: Regression Output (IDE - IMS MAPE Delta)

Distributions

Residual IDE - IMS Delta



Normal

Fitted Normal

Goodness-of-Fit Test

Shapiro-Wilk W

W	Prob<W
0.926974	0.5255

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Figure 96: Residuals Histogram & Shapiro-Wilk Test (IDE - IMS Delta)

Distributions

Studentized Resid IDE - IMS Delta

Quantiles

100.0	maximu	1.29452
99.5%		1.29452
97.5%		1.29452
90.0%		1.29452
75.0%	quartile	0.9818
50.0%	median	0.09588
25.0%	quartile	-1.3398
10.0%		-1.3669
2.5%		-1.3669
0.5%		-1.3669
0.0%	minimum	-1.3669

Summary Statistics

Mean	0.0017571
Std Dev	1.0422086
Std Err Mean	0.3939178
Upper 95% Mea	0.9656394
Lower 95% Mea	-0.962125
N	7

Figure 97: Studentized Residuals Check for Outliers (IDE - IMS Delta)

Table 78: Breusch-Pagan Test for Heteroscedasticity (IDE - IMS Delta)

N	7
Degrees of Freedom model	2
Sum of Squared Errors (SSE)	0.000540
Sum of Squared Residuals (SSR)	1.66E-08
Breusch-Pagan Test Statistic	1.4008
Breusch-Pagan Test p-value	0.4964

Distributions

APE

Quantiles

100.0	maximu	0.9175
99.5%		0.9175
97.5%		0.9175
90.0%		0.9175
75.0%	quartile	0.78846
50.0%	median	0.20431
25.0%	quartile	0.07495
10.0%		0.01653
2.5%		0.01653
0.5%		0.01653
0.0%	minimum	0.01653

Summary Statistics

Mean	0.3697272
Std Dev	0.3636007
Std Err Mean	0.1374282
Upper 95% Mea	0.7060018
Lower 95% Mea	0.0334526
N	7

Figure 98: MAPE for Predicting the Accuracy Delta (IDE - IMS MAPE)

Bibliography

Abdel Azeem, S., Hosny, H. E., & Ibrahim, A. H. (2014, April). Forecasting project schedule performance using probabilistic and deterministic models. *HBRC Journal*, *10*(1), 35-42.

Anbari, F. T. (2003). Earned Value Method and Extensions. *Project Management Journal*, *34*(4), 12-23.

Bolten, J. G. (2008). *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs.* Santa Monica, CA: RAND Corporation.

Bowerman, B. L., & O'Connell, R. T. (1993). *Forecasting and Time Series:An Applied Approach.* Pacific Grove, CA: Duxbury.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control.* Hoboken, NJ: John Wiley & Sons.

Carlberg, C. (2013). *Predictive Analytics: Microsoft Excel.* Indianapolis, Indiana: Que Publishing.

Christensen, D. S. (1993). The Estimate at Completion Problem: A Review of Three Studies. *Project Management Journal*, *24*, 37-42.

Christensen, D. S. (1994). Using Performance Indices to Evaluate the Estimate at Completion. *Journal of Cost Analysis and Management, Spring*, 17-24.

Christensen, D. S. (1999). Using the Earned Value Cost Management Report to Evaluate the Contractor's Estimate at Completion. *Acquisition Review Quarterly, Summer*, 283-296.

Crumrine, K. T. (March 2013, March). *A Comparison of Earned Value Management and Earned Schedule as Schedule Predictor on DoD ACAT I Programs.* MS Thesis, AFIT-ENV-13-M-36, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Defense Acquisition University (DAU). (2014, August 8). Defense Acquisition Guidebook. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=488728

Defense Acquisition University (DAU). (2014, August 8). Gold Card. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=19577

Defense Cost and Resource Center (DCARC). (2014, August 22). Retrieved August 22, 2014, from Defense Cost and Resource Center: http://dcarc.cape.osd.mil/EVM/EVMOverview.aspx

Fleming, Q. W., & Koppelman, J. M. (2000). *Earned Value Project Management.* Newtown Square: Project Management Institute.

Government Accountability Office. (March 2009). *GAO Cost Estimating and Assessment Guide.* Washington DC.

Government Accountability Office. (March 2011). *Defense Acquisitions: Assessments of Selected Weapon Programs.* Washington.

Government Accountability Office. (March 2014). *Defense Acquisitions: Assessments of Selected Weapon Programs.* Washington.

Government Accountability Office. (May 2012). *GAO Schedule Assessment Guide: Best Practices for Project Schedules.* Washington DC.

Henderson, K. (2004). Further Developments in Earned Schedule. *The Measurable News, Spring*, 15-22.

Jack, D. E. (2010). *Contract Over Target Baseline (OTB) Effect on Earned Value Management's Cost Performance Index (CPI).* Wright-Patterson AFB, Ohio: MS Thesis, AFIT-GCA-ENC-10-02. Air Force Institute of Technology.

JMP. (2012). *JMP 10 Basic Analysis and Graphing.* Cary, NC: SAS Institute Inc.

JMP. (2013). *JMP 11 Specialized Models.* Cary, NC: SAS Institute Inc.

Kalman, R. E. (1960, March). A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME—Journal of Basic Engineering*, 35-45.

Keaton, C. (2014). Using Budgeted Cost of Work Performed to Predict Estimates At Completion for Mid-Acquisition Space Programs. *unpublished*.

Kim, B. C. (2007). *Forecasting Project Progress and Early Warning of Project Overruns with Probabilistic Methods.* College Station, Texas: Texas A&M University.

Kim, B. C., & Reinschmidt, K. F. (2010). Probabilistic Forecasting of Project Duration Using Kalman. *Journal of Construction Engineering and Management, 136*(8), 834–843.

Lipke, W. (2003, March). Schedule is Different. *The Measurable News*. Retrieved from http://www.earnedschedule.com/Docs/Schedule%20is%20Different.pdf

Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of Project Outcome: The Application of Statistical Methods to Earned Value Management and Earned Schedule Performance Indexes. *International Journal of Project Management, 27*, 400-407.

Lofgren, E. (2014, June 10). Trust but Verify: An Improved Estimating Technique Using the Integrated Master Schedule IMS. Retrieved from ICEAA Online: http://www.iceaaonline.com/ready/wp-content/uploads/2014/06/EV-3-Paper-Trust-but-Verify-%E2%80%93-An-Improved-Estimating-Technique-Using-the-Integrated-Master-Schedule-IMS.pdf

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications.* New York, NY: John Wiley & Sons Inc.

Montgomery, D. C., Johnson, L. A., & Gardiner, J. S. (1990). *Forecasting and Time Series Analysis.* New York: McGraw-Hill, Inc.

NDIA/PMSC. (2012). *Earned Value Management Systems Intent Guide.* Arlington.

Nystrom, T. D. (September 1995). *A Comparative Study of Linear and Nonlinear Estimate at Completion Methods.* MS Thesis, Air Force Institute of Technology, Wright-Patterson AFB OH.

Petter, J. L. (2014). *An Analysis of Stability Properties in Earned Value Management's Cost Performance Index and Earned Schedule's Schedule Perfomance Index.* Wright-Patterson AFB, Ohio: MS Thesis, AFIT-ENV-14-M-49. Air Force Institute of Technology.

Riposo, J., McKernan, M., & Kaihoi Duran, C. (2014). *Prolonged Cycle Times and Schedule Growth in Defense Acquisition.* Santa Monica, CA: RAND.

Simon, D. (2006). *Optimal State Estimation: Kalman, H, and Nonlinear Approaches.* Hoboken, NJ: John Wiley & Sons.

Smoker, R. E. (2011). Use of Earned Value Management Trends to Forecast Cost Risks. *Journal of Cost Analysis and Parametrics, 4*(1), 31-51.

Trahan, E. (March 2009). *An Evaluation of Growth Models as Predictive Tools for Estimates at Completion (EAC).* MS Thesis, AFIT/GFA/ENC/09-01, Air Force Institute of Technology, Wright Patterson AFB OH.

Unger, E. J. (March 2001). *Relating Initial Budget to Program Growth with Rayleigh and Weibull Models.* MS Thesis, Air Force Institute of Technology, Wright-Patterson AFB OH.

Vandevoorde, S., & Vanhoucke, M. (2006, May). A Comparison of Different Project Duration Forecasting Methods Using Earned Value Metrics. *International Journal of Project Management, 24*(4), 289-302.

Welch, G., & Bishop, G. (2001). An Introduction to the Kalman Filter. Chapel Hill, NC: University of North Carolina at Chapel Hill.

Younossi, O., Lorell, M. A., Brancato, K., Cook, C. R., Eisman, M., Fox, B., . . . Sollinger, J. M. (2008). *Improving the Cost Estimation of Space Systems: Past Lessons and Future Recommendations.* Santa Monica, CA: RAND.

Vita

First Lieutenant Shedrick M. Bridgeforth is from Lac Du Flambeau, WI. He graduated from Lakeland Union High School in Minocqua, Wisconsin. He entered undergraduate studies at the University of Wisconsin Stevens Point where he graduated with a Bachelor of Science degree in Resource Management in May 2006. He attended Officer Training School and was commissioned an officer in the U.S. Air Force in May 2011. In 2012, he was one of seven chosen to attend the Air Force Institute of Technology's Graduate School of Engineering and Management to pursue a Master of Science degree in Cost Analysis. Upon graduation, he will be assigned to the Air Force Cost Analysis Agency, JB Andrews AFB MD.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* 26-03-2015	**2. REPORT TYPE** Master's Thesis		**3. DATES COVERED** *(From – To)* Sept 2013 – March 2015
4. TITLE AND SUBTITLE Using Earned Value Data to Forecast the Duration of Department of Defense (DoD) Space Acquisition Programs			**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Bridgeforth, Shedrick M., 1st Lt, USAF			**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENV) 2950 Hobson Way, Building 640 WPAFB OH 45433-7765			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENV-MS-15-M-177
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Grant Keaton charles.keaton.1@us.af.mil SAF/FMCS Air Force Cost Analysis Agency 1500 West Perimeter Rd, Suite 3500 JB Andrews NAF, MD 20762			**10. SPONSOR/MONITOR'S ACRONYM(S)** AFCAA **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
12. DISTRIBUTION/AVAILABILITY STATEMENT **DISTRUBTION STATEMENT A.** APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.			
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.			
14. ABSTRACT The accuracy of cost estimates is vital during this era of budget constraints. A key component of this accuracy is regularly updating the cost estimate at completion (EAC). A 2014 study by the Air Force Cost Analysis Agency (AFCAA) improved the accuracy of the cost estimate at completion (EAC) for space system contracts. The study found schedule duration to be a cost driver, but assumed the underlying duration estimate was accurate. This research attempts to improve the accuracy of the duration estimate from the AFCAA study; accuracy is evaluated with the Mean Absolute Percent Error (MAPE). The methods researched here are more accurate, timely, and reliable than the status quo method. The original objective, to improve the accuracy of the duration estimates for the cost estimating model, was achieved. The accuracy gains ranged from 2.0% to 13.4% for single contracts, 3.2% to 5.1% for OTB contracts, and 2.9% to 5.2% for all contracts combined. The accuracy improvement is more pronounced from 0% to 70% completion, with a 4.0% to 7.6% increase in accuracy. The accuracy improvement for the EAC was 6.5% (24.4% vs. 17.9%).			
15. SUBJECT TERMS Earned Value Management; Earned Schedule; Schedule; Forecasting; Estimate at Completion			

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Lt Col Jonathan D. Ritschel, AFIT/ENV
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	207	(937) 255-3636, ext 4441 (Jonathan.Ritschel@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18